DOCUMENT RESUME

ED 464 157 UD 034 882

AUTHOR	Metcalf, Kim K.
TITLE	Cleveland Scholarship Program Evaluation: 1998-2000 Technical Report.
INSTITUTION	Indiana Univ., Bloomington. Indiana Center for Evaluation.
PUB DATE	2001-09-00
NOTE	81p.; Appended is "Evaluation of the Cleveland Scholarship Program: 1998-2000 Summary Report." Some figures may not reproduce well.
AVAILABLE FROM	Indiana Center for Evaluation, Smith Research Center, Suite 100, Indiana University, 2805 East 10th Street, Bloomington, IN 47408-2601. Tel: 812-855-4438; Tel: 800-511-6575 (Toll Free); Fax: 812-856-5890; e-mail: iuice@indiana.edu. For full Text: http://www.indiana.edu/~iuice.
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Research (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Educational Vouchers; Elementary Secondary Education; Private Schools; *Program Effectiveness; Public Schools; Scholarships; Student Characteristics; Teacher Characteristics
IDENTIFIERS	Cleveland Scholarship and Tutoring Grant Prog OH

ABSTRACT

This study examined characteristics of students who participated in the Cleveland Scholarship Program and how they compared with nonparticipants; characteristics of the classrooms and teachers of scholarship students and how they compared with those of public school students; and the impact of program participation on students' academic achievement. Data were collected longitudinally. This report examines the second phase of the evaluation, which began in 1998 and focused on children who entered the program (or their public school) as kindergartners or first graders. Researchers collected data on academic achievement, demographics, scholarship utilization, and teacher experience (from classroom interviews). Results indicated that in first grade, scholarship and applicant/non-receipt students were nearly identical demographically, but by late second grade, scholarship students were less likely to be of minority status and were of significantly higher income than were applicant/non-recipients. Classroom and teacher characteristics were relatively similar for public and private school students. Public school teachers were more likely to have completed graduate coursework. There was no clear or consistent academic achievement pattern that could be attributed to participation in the Scholarship Program. By the end of first grade, earlier group differences were reduced or eliminated. (Contains 8 figures and 26 tables.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

THIS DOCUMENT CONCERNS THE CLEVELAND SCHOLARSHIP AND TUTORING PROGRAM AS ESTABLISHED PURSUANT TO OHIO REVISED CODE SECTION 3313.975, AND SHOULD NOT BE CONFUSED WITH THE CLEVELAND SCHOLARSHIP PROGRAM, INC.

ED 464 157

Cleveland Scholarship Program Evaluation

1998 - 2000 Technical Report

Kim K. Metcalf, Ph.D., Director

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

K.K. Metcalf
Indiana Center for Evaluation

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

September, 2001

Indiana Center for Evaluation

Smith Research, Suite 100
2805 East 10th Street
Bloomington, Indiana 47408-2601
http://www.indiana.edu/~iuice/

Kim K. Metcalf, Ph.D.
Director

812-855-4438
800-511-6575
Fax: 812-856-5890
iuice@indiana.edu

UD034882

BEST COPY AVAILABLE

Table of Contents

1 Introduction 1

2 Previous Evaluation Activities and Results 5

2.1 Phase Two of the Cleveland Scholarship Evaluation 10

3 Evaluation Methods and Approaches (1998-2000) 11

3.1 Data Sources and Collection 12
3.1.1 Achievement Test Scores 12
3.1.2 Public School and Scholarship Office Records 13
3.1.3 Classroom Questionnaires 17
3.2 Sample and Sample Selection 18
3.3 Data Analysis Techniques 20

4 Results and Findings 23

4.1 Question One 23
4.1.1 Autumn, 1998 (early first grade) 23
4.1.2 Spring, 1999 (late first grade) 26
4.1.3 Spring, 2000 (late second grade) 28
4.1.4 Summary of Findings Related to Question One 30
4.1.5 Differential Entry of Scholarship Students 30
4.2 Question Two 35
4.2.1 Autumn, 1998 (early first grade) 35
4.2.2 Spring, 1999 (late first grade) 36
4.2.3 Spring, 2000 (late second grade) 38
4.2.4 Summary of Findings for Question Two 39
4.3 Question Three 39
4.3.1 Reading Achievement 40
4.3.2 Language Achievement 43
4.3.3 Mathematics Achievement 45
4.3.4 Total Achievement 48
4.3.5 Summary of Findings Related to Question Three 50

5 Conclusions 51

5.1 Question One 52
5.2 Question Two 54
5.3 Question Three 56
5.4 Summary 60

6 Future Evaluation Activities 61

6.1 Closing Comments 62

List of Tables

TABLE 1. Data Types and Sources 17
TABLE 2. Target and Actual Samples by Testing Episode 19
TABLE 3. Student Demographic Data: Early First Grade, Autumn, 1998 24
TABLE 4. Extended Student Ethnic Data (by percent): Early First Grade, Autumn, 1998 25
TABLE 5. Student Demographic Data: Late First Grade, Spring, 1999 26
TABLE 6. Extended Student Ethnic Data (by percent): Late First Grade, Spring, 1999 27
TABLE 7. Student Demographic Data: Late Second Grade, Spring, 2000 28
TABLE 8. Extended Student Ethnic Data (by percent): Late Second Grade, Spring, 2000 29
TABLE 9. Demographic Data for Scholarship Students of Differential Entry 31
TABLE 10. Fall, 1998 Achievement Data for Differential Entry 33
TABLE 11. Spring, 1999 Achievement Data for Differential Entry 34
TABLE 12. Spring, 2000 Achievement Data for Differential Entry 34
TABLE 13. Classroom and Teacher Characteristics, 1998 35
TABLE 14. Highest Teacher Education Level (in percent) 36
TABLE 15. Classroom and Teacher Characteristics, 1999 37
TABLE 16. Highest Teacher Education Level (in percent) 37
TABLE 17. Classroom and Teacher Characteristics, 2000 38
TABLE 18. Highest Teacher Education Level (in percentages) 39
TABLE 19. Student Achievement in Reading 41
TABLE 20. Mixed Model Analysis of Variance on Students' Reading Achievement Scores: Early First Grade through Late Second Grade 41
TABLE 21. Student Achievement in Language: Early First Grade through Late Second Grade 43
TABLE 22. Mixed Model Analysis of Variance on Students' Language Achievement Scores: Early First Grade through Late Second Grade 44
TABLE 23. Student Achievement in Mathematics: Early First Grade through Late Second Grade 46
TABLE 24. Mixed Model Analysis of Variance on Students' Mathematics Achievement Scores: Early First Grade through Late Second Grade 46
TABLE 25. Student Achievement Total Score: Early First Grade through Late Second Grade 48
TABLE 26. Mixed Model Analysis of Variance on Students' Total Scores: Early First Grade through Late Second Grade 49

List of Figures

FIGURE 1. Minority Status Early First Grade 25
FIGURE 2. Minority Status Late First Grade 27
FIGURE 3. Minority Status Late Second Grade 29
FIGURE 4. Minority Status Differential Scholarship Entry 32
FIGURE 5. Reading Achievement: 1998-00 42
FIGURE 6. Language Achievement: 1998-00 45
FIGURE 7. Mathematics Achievement: 1998-00 47
FIGURE 8. Total Achievement: 1998-00 50

Indiana Center for Evaluation

1 INTRODUCTION

Publicly-funded vouchers have become the education issue of the new millennium. Debate over the value or appropriateness of this idea has been muted though ongoing since Milton Friedman's suggestion of such programs in 1955,[1] but it became much more visible and contentious in the early 1990's when the first publicly-funded voucher program in the U.S. was initiated in Milwaukee. Interest and argument grew over the next decade as Cleveland, Ohio and then the State of Florida implemented similar programs. However, the visibility of the issue reached its highest point yet when then candidate George W. Bush suggested that the idea might be supported at the federal level.

The most recent proposal for a federally funded voucher program was recently rejected by the Congress, and the likelihood of a large-scale federal voucher program, at least in the near future, remains small.[2] However, the issue of publicly-funded school choice that includes private, even religious schools, is larger than ever. More and more states have legislation that supports private school choice or have such legislation pending, and nearly every educational periodical has one or more articles about

1. M. Friedman, *Capitalism and Freedom* (Chicago: University of Chicago Press, 1955).
2. See E. W. Robelen, "Education Bill Ready to Face Final Hurdles" *Education Week* 20, 41 (June 20, 2001): 1, 34-35.

new or proposed choice programs[3]. Despite this intense interest, there is a serious lack of research on the impact and operation of such

programs. The work of Witte and others[4] regarding the Milwaukee voucher program is now nearly a decade old and, because Wisconsin discontinued evaluation of the program, no new research is likely to be generated from the longest running state-funded voucher program in the country. Florida's statewide voucher program serves only a very small number of students (fewer than 100 during the 1999-00 academic year) and includes no independent evaluation component. It is in this regard that the Cleveland Scholarship Program represents such an important and visible example in the current debate. The targeted nature of the Cleveland Scholarship Program, its longevity, and the integrated but independent evaluation it includes make the activities and experience of Cleveland vital to the national debate on school choice.

The present report reflects the most recent progress and findings associated with one of these elements: ongoing evaluation of the program's effects. The report is the fourth since the program began, and describes the methods and results of data drawn over a two-year period on students who entered public school or the Scholarship Program as kindergartners in fall, 1997, first graders in fall 1998, or second graders in fall 1999. The primary focus of evaluation during this period was on the characteristics of these students, the characteristics of their classrooms and teachers, and the academic progress of the students from early first grade through the end of second grade. The data and results represent the two-year findings of the ongoing longitudinal evaluation.

3. In addition to voucher programs, which directly support private school enrollment, a number of states have in place tuition tax credit programs that allow families to deduct a portion of private school tuition. Further, a majority of states have approved charter school legislation that often includes religiously affiliated schools to participate.
4. See John F. Witte, *The Market Approach to Education: An Analysis of America's First Voucher Program* (Princeton, NJ: Princeton University Press, 2000) for a comprehensive presentation of the entirety of this work.

Fundamentally, the evaluation seeks to provide empirically-derived information on the operation of the program; its impact on students, families, schools, and educators; and the implications of the Cleveland program for voucher programs in other areas. In each year of the project, specific evaluation questions have been emphasized. In the period covered by the present report, the following specific questions were addressed:

1. What are the characteristics of students who participate in the Scholarship Program, and how do they compare with students who do not participate?
2. What are the characteristics of the classrooms and teachers experienced by scholarship students and how do they compare with those experienced by public school students?
3. What is the impact of participation in the program on students' academic achievement?

The report is organized into four basic sections. In a first section, preceding evaluation activities (those conducted from 1997-99) are briefly reviewed. A second section describes the evaluation approaches and methodologies that were used during the period of this report. A third section presents the results of data analysis and is organized around the three evaluation questions. A fourth section describes the current results in the context of earlier evaluation findings. The report ends with a brief summary that explains the intended future activities of the evaluation.

Indiana Center for Evaluation

2 PREVIOUS EVALUATION ACTIVITIES AND RESULTS

Evaluation of the Cleveland Scholarship Program was initiated by the Indiana Center for Evaluation in the spring of 1997, at the conclusion of the first operating year of the program. From the beginning, the evaluation has examined the impact of the Scholarship Program on the students, families, and schools that participate, and the impact on the public schools from which the scholarship students are drawn. In the first three years of the program, the evaluation focused primarily on 183 scholarship students who entered the program as third graders during the 1996-97 academic year. From spring 1997 through spring 1999, the academic achievement of these students was measured annually by the evaluation team. In addition, data were collected from public school and scholarship office records, telephone and written surveys, and classroom interviews. These supplemental data were used both for descriptive purposes (i.e., to understand the children, families, and schools) and to make analyses of achievement more efficient.

Detailed findings from this early stage of the evaluation have been reported elsewhere.[5] However, the results indicated that:

- Scholarship students had been achieving at higher academic levels than their classmates as second graders in public schools, prior to entering the Scholarship Program.

5. See K. Metcalf, W. Boone, P. Muller, P. Tait, and N. Stacey, *Evaluation of the Cleveland Scholarship Program: 1997-1999* (Bloomington, Indiana: Indiana Center for Evaluation, 1999).

- Scholarship students were significantly more likely to be from families of low income, headed by a single mother, and African-American than their public school counterparts.
- Classrooms attended by scholarship students had significantly fewer students, and were headed by teachers with significantly less experience and who were less likely to have completed graduate coursework than those attended by public school students.
- Parents of scholarship students were more satisfied with their children's school and classmates, were more likely to be involved with the school, and had been less satisfied with their children's former (public) school than parents of public school students who had not applied for a scholarship.
- Teachers and administrators in public and private schools had mixed feelings about the program, believing it had both positive and negative impacts on their schools. Private school educators felt that the program had initially diminished the level of academic achievement in their classes and had made their jobs more difficult, but they also felt that the scholarship students added valuable diversity to their schools. Public school educators felt that the program took only the "best and brightest" students and most involved families and that it reduced support for public education, but they also felt that the small number of students involved made little difference for them or their school.
- After controlling for initial differences in academic achievement and a limited set of demographic and classroom factors, scholarship students achieved at significantly higher levels than their public school counterparts in language and science. However, this was true only for students who attended private schools that existed prior to the Scholarship Program and did not apply to students attending private schools established solely to serve scholarship students.

In discussing these findings, and when placing them in the context of other research of voucher programs, we noted that they are relatively consistent with patterns found by other researchers. Most studies have found that voucher programs, whether publicly or privately-funded, tend to promote more positive parental or family attitudes toward school, increase parental involvement, and result in increased parental satisfaction.[6] In addition, the limited, but statistically significant positive impact of the program on student's academic achievement, particularly as they progress beyond the early primary grades, is consistent with work in Milwaukee, New York, Washington, DC., and

6. See Witte, J. P. Greene, P. E. Peterson, and J. Du "School Choice in Milwaukee: A Randomized Experiment" (Cambridge, MA: Program on Education Policy and Governance, Harvard University, 1997).

Dayton, Ohio.[7] Our early work in Cleveland, however, provided unique findings related to teacher and administrator perceptions of the program's impact on schools and on the reasons for parental selection of schools.

An important outgrowth of this early work was associated with the most appropriate methods by which to examine the impact of voucher programs. Considerable debate has been generated over how best to determine whether voucher programs affect student achievement. In Milwaukee, as in the first three years of our evaluation in Cleveland, participating students were compared to students who continued to attend public schools and whose families had never sought a voucher for private school enrollment. In these quasi-experimental approaches, the comparison group cannot be assumed to be completely similar to the voucher group; thus, comparisons may be somewhat flawed. For example, it is known from the outset that parents who seek a voucher for their children differ from those who do not at least in their motivation or willingness to pursue a voucher and, perhaps, in their dissatisfaction with public schools. It is reasonable to believe that these differences, and others that we may not know of, will influence a child's performance and achievement, no matter what school she or he attends. In these studies, statistical techniques (e.g., analysis of covariance are used to control for known differences between groups, but they still may not provide completely equivalent comparison groups. We, and the researchers who conducted the original studies in Milwaukee, contend that such approaches are appropriate and, in these cases, were necessary because no "pure" comparison group existed.[8]

Others, however, contend that the most appropriate comparison group is a randomly chosen group of families and children who applied for but were not selected in the lot-

7. See Witte; W. G Howell, P. J. Wolf, P. E. Peterson, and D. E. Campbell, "Test-Score Effects of School Vouchers in Dayton, Ohio, New York City, and Washington, DC: Evidence from Randomized Field Trials" (Paper presented at the annual meeting of the American Political Science Association, Washington, DC, September, 2000).

8. In the early years of the Cleveland and Milwaukee programs, all voucher applicants were offered a voucher. As a result, there were no students or families who had wanted a voucher (and thus were similar to other voucher applicant families) who did not have the opportunity to use one.

tery process to receive a voucher.[9] These families, they believe, offer a group that is most like those who participate (in that they were motivated to pursue a voucher and were, seemingly, dissatisfied with public schools). In these cases, an experimental approach is possible that compares the impact of the program on these two, randomly established (through the voucher selection lottery process) groups. In evaluating privately-funded programs, such as those in New York and Washington, DC, such an experimental approach has been used. However, it is possible that this approach may still not provide comparison groups that are equivalent in critical aspects.

Such random selection of awardees may, in fact, produce two similar groups at the outset. All of those in the original sample, those selected and those not selected, were motivated to pursue a voucher for private school enrollment. However, because all families in both groups wished to use a voucher to send their children to private schools, but only those who received a voucher were able to do so, non-selected families are essentially forced to send their children to schools that they wished to leave. It is reasonable that this may, in fact, have a negative effect not only parental satisfaction with schools, but may also affect children's achievement. Thus, even random selection, a hallmark of clinical and experimental research in education, cannot in itself provide the most complete comparisons.

Our work in Cleveland suggests that each of these approaches has both merits and problems that are likely to impact the findings of research on vouchers. Most importantly the work indicates that no single comparison group is sufficient. For all of the factors noted above, the evaluation in Cleveland suggests that a combination of comparison groups, while logistically more difficult, provides the most complete information about program effectiveness. Whenever possible, voucher participants should be compared to at least two other groups of non-participants: (1) families and children

9. Howell et al.; W. G. Howell, P. J. Wolf, P. E. Peterson, and D. E. Campbell, "The Effect of School Vouchers on Student Achievement: A Response to Critics," (Occasional paper, Program on Education Policy and Governance, Harvard University, 2001).

who applied for a voucher, but who were not selected to participate and, thus, who continue to attend public schools; and (2) families and children who did not apply for a voucher and who continue to attend public school. The first group provides a comparison with families who were both motivated to seek out a voucher for their children and who were dissatisfied with public schools. However, they are also subject to the bias that results from their non-selection. The second group represents families who may differ from voucher applicants in their interest or motivation to seek out a voucher for their children, but they may also be more satisfied with the public schools their children attend. Our work in Cleveland suggests that, whenever possible, at least these two groups should be used in making comparisons.

At least one other group is also appropriately included, when sufficient in size and accessibility. This is the group of families who applied for and were offered a voucher (thus they are motivated and have some sense of dissatisfaction with public schools), but who elected not to use the voucher and, instead, continue to enroll their children in public schools. Unfortunately, this group is often quite small in number and difficult to locate.

The work described above focused, as noted earlier, on a cohort of students who entered the Scholarship Program as third graders in 1996-97, and who had attended public schools as second graders the preceding year. Data on these students and their families were supplemented by collection of data from randomly selected samples of parents/guardians, and school personnel across the grades served by the program. The first phase of the evaluation concluded with release of our comprehensive report in autumn, 1999. However, a second phase of the evaluation was, by then, already underway.

2.1 Phase Two of the Cleveland Scholarship Evaluation

The second phase of the evaluation, which began in autumn, 1998, is initially reported in the present document, and is intended to address issues or questions that could not be answered in phase one. In particular, the second phase was conceptualized to: (a) focus on a cohort of children who entered the program (or their public school) as kindergartners or first graders, (b) use multiple comparison groups (as described above), and (c) to examine longitudinally the impact of the program on children, families, and schools. As a result, phase two was initiated in autumn, 1998 when achievement data were first collected on 780 first grade scholarship students, 541 first grade public school students whose parents had applied for a voucher but had not received one, and 1,233 first grade public school students whose parents had not applied for a voucher. The remainder of this report details the evaluation activities, findings, and implications of the initial two years of this second phase of the evaluation.

3 EVALUATION METHODS AND APPROACHES (1998-2000)

As noted above, phase two of the evaluation was designed to expand information gained earlier and to examine the impacts of school choice longitudinally, following a cohort of children across several years of participation. Over a multi-year period, this phase is expected to provide information associated with the real and perceived impact of the program on schools, teachers, families, and children. However, the first two years of this phase were focused on identifying the target sample of students (participating and non-participating) who would constitute the sample and on collecting baseline demographic and achievement data on these children.

Guided by the research questions presented earlier in the report, the evaluation uses a mixed-model research design and multiple comparison groups. The fundamental source of data for this phase is the Terra Nova (CTB/McGraw-Hill), an achievement test independently administered by representatives of the evaluation team. In addition, data are collected from Cleveland Public School and Cleveland Scholarship Office records, teacher interviews, and limited classroom observations by evaluation team proctors. In future years, data will be collected from representative samples of public and private school families, teachers, and other school personnel; from state and district records associated with resource allocation; and from focused interviews of upper level school representatives, department of education staff, and other stakeholders.

3.1 Data Sources and Collection

As noted above, data were drawn from three primary sources and one ancillary data source for the period covered by this report. The fundamental source of data was the Terra Nova, used as a measure of student's academic achievement. Supplementing these achievement data were those drawn from Scholarship Office and Cleveland Public School records. A third data source was a brief interview/observation record of teacher and classroom characteristics that was made by evaluation team proctors at the time of testing. In addition, data reflecting student mobility during the two-year period were available through records maintained by the evaluation team. Each of these is described in detail below.

3.1.1 Achievement Test Scores

Student's academic achievement throughout phase two of the evaluation is being measured using the Terra Nova (CTB, McGraw-Hill). This achievement test was selected early in the project because: none of the schools in which data are collected use the Terra Nova as their primary off-grade testing tool, the test is among the most progressive in its use of contemporary testing principles and formats, and scores can be provided that allow linear comparisons to be made across multiple grade levels of the test. While the scope of the test expands in upper elementary grades, in first and second grades it provides four scores for each student: reading, language, mathematics, and total score.

In order to collect baseline achievement data, the Terra Nova, Level 10 was administered to first grade students in the study during November, 1998. It was believed that testing in November ensured that most children were acclimated to their schools and to school attendance. Level 10 of the Terra Nova was selected because it was appropriate for kindergarten and early first grade children. Using standardized administration procedures, proctors who were trained and supervised by the evaluation team

administered the test in two sessions to groups of target students in public and private schools. The first session of testing in each classroom was conducted in a brief afternoon session and was used to complete the reading and language portions of the test. A second session was conducted the following morning in each classroom and allowed students to complete the mathematics portion of the test. Two adult proctors were assigned to each group of students to be tested to provide sufficient supervision and assistance to the young children.[10]

In spring, 1999, attempts were made to test these targeted students using Level 11 of the Terra Nova. As before, the test was administered by trained proctors using a standardized format that included an afternoon and following morning session; and again two proctors were assigned to each group of students. This same process was used to administer the Terra Nova, Level 12, in April, 2000, when the target students were nearing the end of their second grade year.

The longitudinal nature of the current project requires the use of linear scaled scores, rather than the more commonly used but non-linear normal curve equivalent (NCE) scores. Scaled scores on the Terra Nova range from 0-1000 and, unlike non-linear scores (e.g., grade equivalent or percentile rank scores), can be compared and combined across multiple years of testing.

3.1.2 Public School and Scholarship Office Records

The primary source of existing data was Cleveland Scholarship Program records. These records were used in two ways. First, records provided information that allowed identification of students whose families had applied for a scholarship, whether or not a scholarship was offered, whether or not an offered scholarship was

10. Achievement data were to be used solely for comparative purposes among participating and non-participating students. As a result, while afternoon testing is usually not recommended., because all students in all groups took the test under the same conditions (i.e., one afternoon and one morning session), it was considered appropriate

BEST COPY AVAILABLE

accepted and used, and the private school in which scholarship students were enrolled. Working with representatives of this office, records were used to determine what would become the core sample of students to be targeted for data collection during this phase of the evaluation. These students represented: (a) those who had applied for, accepted, and were using a scholarship to attend a private school; (b) those who had applied for, but not been awarded a scholarship and who attended public school[11]; and (c) those who applied for, accepted, and had used a scholarship in the past, but who had returned to public schools.[12] These groups will be further described in the Sample section of this report.

A second use of Scholarship Office records was to obtain demographic information on students whose families had applied for a scholarship at some point during their kindergarten, first, or second-grade years. This information was used not only to provide descriptive information about the students and their families, but it also allowed comparisons of participating and non-participating students in later analyses. Scholarship Office records are updated at least annually and were used to obtain data on: student gender, student race, family size, and family income.

Records maintained by Cleveland Public Schools were also used in the present study and in similar ways to that of Scholarship Office records. Initially, these records were used to locate the school of enrollment for students whose families had applied for a scholarship but who had either not received or were not using the scholarship to attend private school. These were students in groups (b) and (c) above.

Cleveland Public School records were also used to obtain demographic information on students attending public schools. This information included: student race, student

11. In real terms, this group consisted of two types of students: those who were not selected in the random lottery to award scholarships, and those whose family income was above the federal poverty level and, as a result, they were given low priority in the selection process and may not have been included in the lottery.
12. In the present sample, this group consisted of only 18 students by the end of second grade and, thus, was too small to be included in most analyses. However, when possible, supplemental analyses were conducted and are reported to describe this group of students.

BEST COPY AVAILABLE

gender, and eligibility for free or reduced price lunch. For students whose families had applied for a scholarship, these data were used to validate information obtained from the Scholarship Office. For students whose families had not applied for a scholarship, this was the only demographic information available.

Family income has often been found to be strongly related to students' academic achievement and, thus, was included as one variable in our earlier evaluation work. It was believed important to include some estimate of this variable in the present phase of the evaluation. However, no single measure of family income was available for all students. Scholarship students who had never attended public school (most of the present participating sample) had provided the Scholarship Office with family income data. Public school students whose families had never applied for a scholarship could be identified as eligible for free or reduced price lunch, but there was no linearly scaled measure of family income available. Importantly, however, 118 children existed whose families had applied for a scholarship (and thus had provided a measure of family income), but who were or had attended public schools (and thus had been assigned a code for their eligibility for free or reduced price lunch).

In order to attempt to calculate a valid, consistent estimate of family income, the two available forms of income data on these 118 children were used to calculate an estimated meal code. The process by which this was accomplished was as follows:

1. Using assigned meal code (i.e., eligibility for free and reduced price lunch reported in Cleveland Public School records) to classify students, mean per capita family income, and 95% confidence intervals were computed for each meal code (free lunch, reduced price lunch, or paid lunch).
2. Confidence intervals were then used to define new estimates of meal code by applying them to available linear measures of per capita family income. The resulting codes ranged from 1 for per capita family income up to $3,684; 2 for income ranging from $3,685 to $6,580; and 3 for income above $6,580.
3. A new variable, estimated meal code, was created and reflected either: (i) assigned meal code from Cleveland Public Schools or (ii) the newly created estimate of meal code calculated above.

This process allowed definition of an estimated meal code for nearly all students in the sample, whether or not they had attended public school or had applied for a scholarship. It should be noted that these estimated meal codes are not exact measures of family income. It is likely that they provide a somewhat inflated indication of family income when compared with meal codes assigned by the public schools. The reason for this is that all students in public school are categorized as eligible for free lunch if 60% or more of the students in the school are eligible based on income. Nonetheless, because an estimate of income is important in the present study, and because the relationship between this estimated meal code and other measures of income (i.e., assigned meal code and family income) is strong, it is used as a proxy indicator of family income.[13]

An ancillary source of demographic data in the present project was information maintained by the evaluation team over the course of the work. Evaluation records indicate a numeric code representing each student's school of enrollment at each testing episode. As a result, for many students, it was possible to develop a basic indicator of student mobility across schools over the two-year period.

As described above, there was substantial inconsistency in the availability of student data, particularly across the differing record keeping systems of the Scholarship Office and Cleveland Public Schools. In spite of this, at least some demographic data were available on all students in the sample. These variables included:

- Student gender: coded 1 = Male, 2 = Female
- Student race[14]: coded 1 = Non-minority, 2 = Minority
- Estimated meal code: coded 1, 2, or 3 as described above

13. Simple correlations between this estimated meal code and assigned meal code or reported family income are r=.73 and r=.78, respectively.

14. Scholarship Office and Cleveland Public School records list race as: African-American, Caucasian, Hispanic, Multiracial, and other. Due to small numbers of students in all but the African-American and Caucasian classifications, these were recoded dichotomously as minority and non-minority (i.e., Caucasian).

- Student mobility: coded 0 = no school changes, 1 = one school change, 2 = at least two school changes

3.1.3 Classroom Questionnaires

The third principle source of data for the current project was a brief classroom interview of each teacher conducted by the proctors on the days they tested in each school. Proctors used a guided questionnaire to obtain information about the teachers with whom the public and private school students worked and about the size of the classes in which they were enrolled. Specifically, proctors spoke with each teacher whose students were included in the testing to obtain data on the teacher's: (a) class size, (b) years of experience, and (c) level of education. The principle purpose of these data was to describe, at least superficially, the classroom context in which the students worked.

Each type of data and the source from which it was obtained is presented in Table 1 below. The data described above were used independently and in combination in all subsequent analyses.

TABLE 1. Data Types and Sources

	Source of Data				
Type of Data	**Public School Records**	**Scholarship Office Records**	**Terra Nova**	**Evaluation Office Records**	**Classroom Interviews**
Student Gender	✔	✔			
Student Race	✔	✔			
Family Size		✔			
Family Income		✔			
Meal Code	✔				
Estimated Meal Code				✔	
Academic Achievement			✔		
School of Enrollment			✔		
Class Size					✔
Teacher Experience					✔
Teacher Education					✔

3.2 Sample and Sample Selection

As noted earlier, research and evaluation of voucher programs has generated substantial debate over the most appropriate comparison group against which to judge program impact. In the present study, we attempted to include multiple comparison groups to ensure that the most complete and representative judgments could be made. In order to initiate the longitudinal evaluation, an early goal was to develop the broadest possible sample of participating and non-participating students during their first-grade year in 1998-99 and to follow these students over a multi-year period.

From the beginning of this phase, the evaluation intended to identify and follow students in each of four groups. These groups included: (a) scholarship students attending private schools; (b) scholarship applicant/non-recipients, students who applied for but did not receive a scholarship and who attend public schools; (c) non-users, students who received and possibly used a scholarship for one or more years, but who now attend public schools; and (d) non-applicants, public school students whose families never applied for a scholarship.

Students in groups (a), (b), and (c) were first identified through Scholarship Office records. These records also indicated scholarship students' private schools of enrollment. However, school of enrollment was not available for either applicant/non-recipients or scholarship non-users. Public school records and private school contacts were used in an attempt to determine the school of enrollment for students in these groups.

The final comparison group consisted of public school students whose families had not applied for a scholarship. This group of students was identified in the first year of the evaluation as testing was conducted in every public school and all participating private schools in Cleveland. Students in this group represent children who were enrolled in public school classrooms with students identified in groups (b) and (c) above. In order to provide the most comfortable and consistent environment for test-

ing, most schools elected to test all students in classrooms, even if they had not been identified as a target student. As a result, in the process of testing scholarship applicant/non-recipients and scholarship non-users in public schools, a large number of non-applicant public school students were also tested. These students came to constitute the fourth group with which scholarship students were compared.

This process began in autumn, 1998, when students were beginning their first-grade year. It was repeated and updated in spring, 1999 and spring, 2000 as students were tested at the end of first grade and at the end of second grade. Through this process of updating student status, and because students move into and out of the various groups, sample sizes vary across testing episodes. Further, the sample can be viewed as consisting of a target sample (students who were classified through records or earlier testing into one of the four target groups) and an actual sample (students who were classified into one of the four groups and for whom achievement or demographic data were obtained).

Over the period covered by the present report, the target and actual samples are as follows:

TABLE 2. Target and Actual Samples by Testing Episode

Group/Status	Time of Testing					
	Autumn, 1998		Spring, 1999		Spring, 2000	
	Target	Actual	Target	Actual	Target	Actual
Scholarship Recipient/Users	894	780	789	737	695	640
Applicant/Non-Recipients	612	541	600	536	658	499
Non-Users[1]	NA	NA	NA	NA	18	12
Public School Non-Applicants		1233	1224	1094	986	836

1. Data available for the first grade period did not distinguish between *applicant/non-recipients* and *non-users*.

3.3 Data Analysis Techniques

Data analysis was conducted to focus on each of the three evaluation questions guiding this phase of the evaluation and on relevant emerging subquestions. For each question, analyses included both descriptive and inferential techniques. Inferential analyses relied upon analysis of variance approaches and, when appropriate, follow-up pairwise comparisons. Evaluation Question 1 was investigated using descriptive approaches and multivariate analyses of variance for each of the three sets of data resulting from the three testing episodes. When appropriate, univariate analysis of variance and modified Tukey-Kramer[15] procedures for post hoc comparisons were used to further understand indicated differences. In each set of analyses, students were the unit of analysis and the largest possible sample sizes were used.[16] Because the specific students included in each set of analyses are not exactly the same, some caution must be used when attempting to compare across the three episodes.

An additional element of Question 1 was comparison of students who entered the Scholarship Program as kindergartners, first-graders, or second-graders. Descriptive and inferential analyses were conducted on student demographic and achievement data across the three testing episodes to answer this question. It was hoped that similar analyses could be conducted to investigate differences between students who remained in the program and those who chose to return to public school. However, only 12 of 18 students who left the program could be located in public schools. As a result, this comparison was not possible.

15. See J. W. Tukey, *The problem of multiple comparisons* (Princeton University: Ditto, 1953); J.W.Tukey, *Exploratory Data Analysis* (Reading, MA: Addison Wesley, 1977); and C. Y. Kramer, "Extensions of Multiple Range Test to Group Means with Unequal Numbers of Replications," *Biometrics* 12 (1956): 307-310.
16. For example, data were analyzed on all students in each of the comparison groups for whom sufficient data were available in autumn, 1998, even if these same students could not be included in one or more of the following sets of analyses. Similarly, students who had been targeted but not tested in an earlier period, but for whom sufficient data were available later in the project were include in later analyses.

BEST COPY AVAILABLE

Evaluation Question 2 similarly used the largest possible sample for each of three independent sets of analyses. However, unlike analyses of Question 1, which used students as the unit of analysis, Question 2 relied on data aggregated by classroom. Further, because the primary comparison was between public and private schools rather than between or among the three groups of students, the analyses attempt to investigate differences in classroom and teacher characteristics by classroom and between public and private schools. For data from each testing episode, descriptive, multivariate, and univariate analyses of variance were conducted.

Evaluation Question 3, addressing students' academic achievement was investigated somewhat differently than either of the two above questions. In this case, it was believed important not just to examine overall differences in achievement between and among the three student groups but also to focus attention on the pattern of achievement for each group from the beginning of first grade through the end of second grade. As a result, data analyses for Question 3 relied primarily on mixed model analyses of variance.[17] These analyses provide the opportunity to investigate the impact of program participation over time, and were believed to be the most defensible and easily interpreted. Thus, for Question 3, the analyses for each subject measure examined differences in the academic performance of students: (i) between the scholarship, applicant/non-recipient, and non-applicant groups; (ii) between each of the three testing episodes; and, most importantly for the present evaluation, (iii) the interaction of group membership and time (i.e., do students in one group improve at a greater rate than other students).

These analyses were further refined to examine possible differences in achievement between scholarship students who entered the program as kindergartners and scholar-

17. For a discussion of the use and interpretation mixed model analysis of variance (in this case one-between/on-within subjects), see J. Stevens, *Applied Multivariate Statistics for the Social Sciences, 3rd. Edition*, (Mahwah, NJ: Erlbaum Associates, 1996); J. J. Kennedy and A. J. Bush, *An Introduction to the Design and Analysis of Experiments in Behavioral Research*, (Lanham, MD: University Press of America, 1985).

ship students who entered the program in first grade. It had been hoped that these analyses could also examine the group of scholarship students who entered the program in second grade. However, achievement data were available for only 16 students in this group, and all of these students had attended private school during first grade. As a result, the set of four, mixed model analyses of variance each included four groups: 3-year scholarship students (who entered in kindergarten), 2-year scholarship students (who entered in first grade), applicant/non-recipients, and non-applicants. For all analyses, only students for whom data were available for all three testing episodes were included.

Indiana Center for Evaluation

4 RESULTS AND FINDINGS

Results and findings are organized around the three evaluation questions that guided the initial years of phase two of the project. For each question and related subquestions, the results of data analysis are presented and briefly described. Implications of or conclusions from these findings are presented in a subsequent section of the report.

4.1 Question One

What are the characteristics of students who participate in the Cleveland Scholarship Program and how do they compare with students who do not participate?

Analyses of Evaluation Question 1 included description and comparison of the demographic characteristics of students in each of the primary groups, using the largest available sample for each. Three sets of analyses were conducted, one for each testing episode, and they included multivariate analyses of: student gender, student minority status, student mobility, and estimated meal code as a proxy for family income. The results of these analyses are presented by testing episode below.

4.1.1 Autumn, 1998 (early first grade)

Table 3 presents descriptive statistics associated with the demographic characteristics of students in the three primary groups. Because this was the first testing episode and no previous enrollment data were available, student mobility was not included in these analyses. Multivariate analyses supported conducting univariate and pairwise com-

parisons on estimated meal code, minority status, and gender (Pillai's Trace = .018, p = .015). These analyses revealed no significant differences between scholarship students and their non-participating peers in estimated meal code (F [2, 895] = 1.45, p = .310) or gender (F [2, 895] = .27, p = .766).

TABLE 3. Student Demographic Data: *Early First Grade, Autumn, 1998*

		Estimated Meal Code	Minority Status[1]	Gender[2]
Scholarship Recipients N = 557	Mean	1.68	1.30	1.52
	S.D.	.823	.457	.500
Public Applicant/ Non-Recipients N = 174	Mean	1.56	1.21	1.51
	S.D.	.771	.410	.500
Public Non-Applicants N = 167	Mean	1.64	1.17	1.49
	S.D.	.808	.375	.500
Total	Mean	1.65	1.26	1.51
	S.D.	.811	.437	.500

1. For these analyses, Minority Status was coded dichotomously as Non-minority = 1, Minority = 2. As a result, numbers following the decimal indicate the percentage of students who where of minority status.
2. For these analyses, Gender was coded dichotomously as Male = 1, Female = 2. As a result, numbers following the decimal indicate the percentage of students who were female. This interpretation can be applied to all subsequently reported analyses of Gender.

However, there were significant differences among the groups in ethnic status (F [2, 898] = 6.72, p = .001). Table 4 presents expanded minority status data for the groups and Figure 1 presents these data graphically. Follow-up pairwise comparisons.[18] found that non-applicant students were significantly more likely to be from ethnic minorities than were scholarship students. In contrast, there were no differences in ethnic status between scholarship and applicant non-recipients or between applicant/ non-recipients and non-applicants.

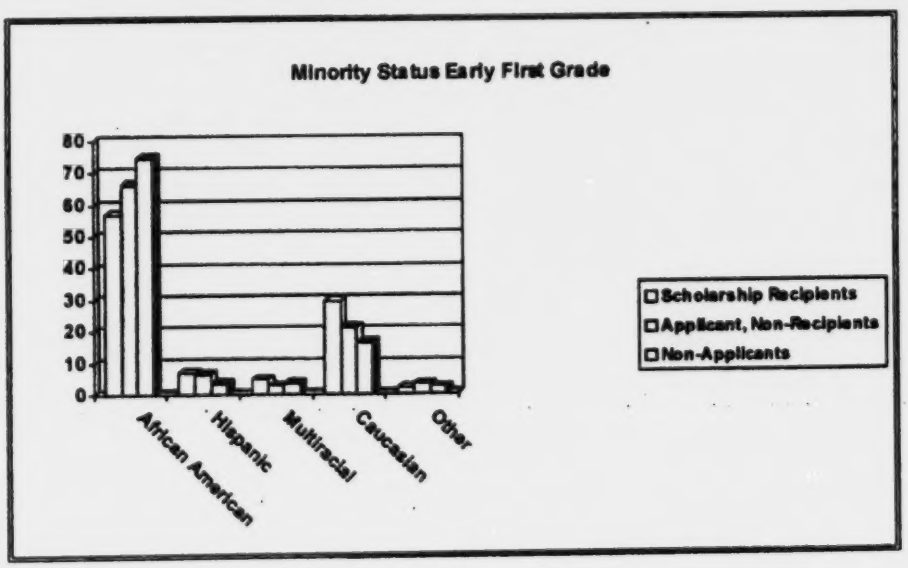
18. All pairwise comparisons throughout this report were conducted using modified Tukey-Kramer techniques to control overall experimentwise error at less than $\alpha < .05$.

BEST COPY AVAILABLE

TABLE 4. Extended Student Ethnic Data (by percent): *Early First Grade, Autumn, 1998*

	African American	Hispanic	Multiracial	Caucasian	Other	Total
Scholarship Recipients	56.8%	7.0%	4.8%	29.5%	1.9%	100%
Public Applicant/ Non-Recipients	66.1%	6.6%	2.7%	21.3%	3.3%	100%
Public Non-Applicants	74.6%	3.4%	3.4%	16.4%	2.3%	100%
Total	61.9%	6.2%	4.1%	25.5%	2.2%	100%

FIGURE 1. Minority Status Early First Grade



4.1.2 Spring, 1999 (late first grade)

Table 5 presents descriptive demographic data for students tested in spring, 1999. Multivariate analyses, including estimated meal code, minority status, gender, and mobility indicated differences among the groups on the four variables (Pillai's Trace = .020, p = .008). Univariate analyses indicated no significant differences among the groups in estimated meal code (F [2, 891] = 1.72, p =.180), gender (F [2, 891 = .37, p = .695), or mobility (F [2, 1149] = .948, p = .388), but suggested significant differences among the groups in minority status (F [2, 891] = 7.31, p = .001).

TABLE 6. Student Demographic Data: *Late First Grade, Spring, 1999*

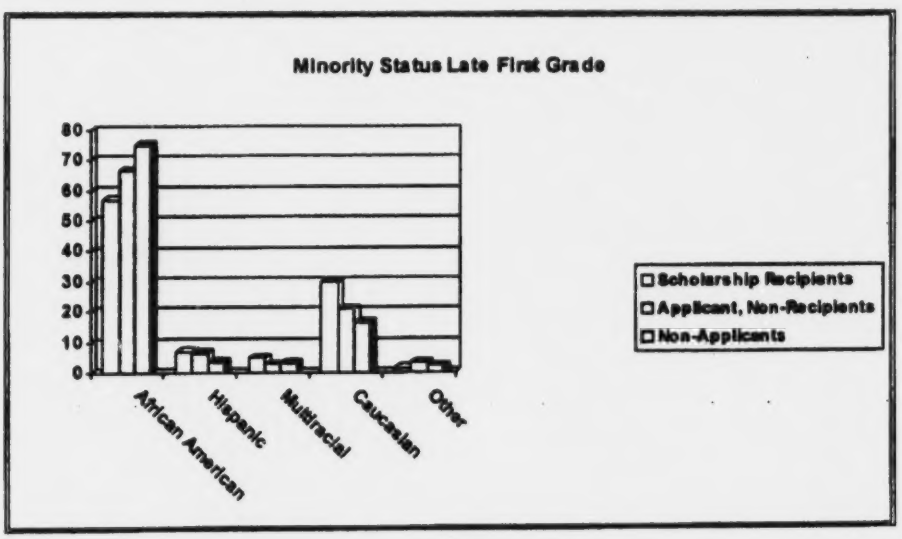
		Estimated Meal Code	Minority Status	Gender	Mobility
Scholarship Recipients N = 557	Mean	1.68	1.30	1.52	.017
	S.D.	.823	.458	.050	.130
Public Applicant, Non-Recipients N = 173	Mean	1.55	1.21	1.50	.037
	S.D.	.765	.407	.050	.190
Public Non-Applicants N = 164	Mean	1.63	1.16	1.48	.028
	S.D.	.807	.372	.050	.170
Total	Mean	1.65	1.26	1.51	.025
	S.D.	.810	.437	.050	.160

Pairwise comparisons of minority status data indicated that non-applicant public school students were significantly more likely to be of minority status than were scholarship students. There were no significant differences in minority status between applicant/non-recipients and non-applicants. These patterns are illustrated in the expanded minority status data presented in Table 6 and Figure 2.

TABLE 6. Extended Student Ethnic Data (by percent): *Late First Grade, Spring, 1999*

	African American	Hispanic	Multiracial	Caucasian	Other	Total
Scholarship Recipients	56.8%	7.0%	4.8%	29.7%	1.7%	100%
Public Applicant, Non-Recipients	66.5%	6.6%	2.7%	20.9%	3.3%	100%
Public Non-Applicants	75.1%	3.5%	2.9%	16.2%	2.3%	100%
Total	62.1%	6.3%	4.0%	25.5%	2.1%	100%

FIGURE 2. Minority Status Late First Grade



4.1.3 Spring, 2000 (late second grade)

Student demographic data were again analyzed for students in each group who were tested in spring of second grade. Table 7 presents descriptive data on estimated meal code, minority status, gender, and mobility.

TABLE 7. Student Demographic Data: *Late Second Grade, Spring, 2000*

		Estimated Meal Code	Minority Status	Gender	Mobility
Scholarship Recipients N = 574	Mean	1.72	1.33	1.51	092
	S.D.	840	493	050	290
Public Applicant, Non-Recipients N = 110	Mean	1.78	1.22	1.44	11
	S.D.	828	415	050	320
Public Non-Applicants N = 296	Mean	1.52	1.15	1.53	15
	S.D.	768	356	050	360
Total	Mean	1.66	1.26	1.51	13
	S.D.	823	455	050	340

On the basis of initial multivariate analysis of variance (Pillai's Trace = .049, $p < .000$), separate univariate analyses were conducted on each of the four variables. These analyses identified no significant differences among the groups on gender ($F [2, 977] = 1.53$, $p = .218$), but revealed significant between group differences on estimated meal code ($F [2, 977] = 7.09$, $p = .001$), minority status ($F [2, 977] = 17.18$, $p < .000$), and mobility during first and second grade ($F [2, 1156] = 4.22$, $p = .015$).

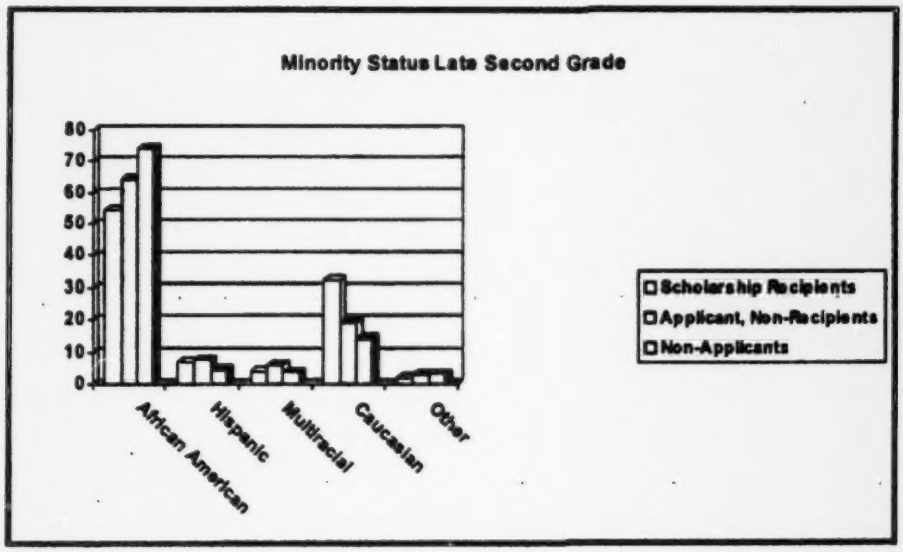
Follow-up pairwise comparisons were conducted to explicate significant differences. Scholarship students were found to have significantly higher estimated meal codes than non-applicant public school students; however, there was no significant difference in estimated meal code between scholarship recipients and applicant/non-recipients. Scholarship students also were much less likely to be of minority status than either applicant/non-recipients or non-applicants. This pattern is reflected in Table 8

and Figure 3. In addition, scholarship students changed schools significantly less often than non-applicant public school students. However, there was no significant difference between applicant/non-recipients and non-applicants

TABLE 8. Extended Student Ethnic Data (by percent): *Late Second Grade, Spring, 2000*

	African American	Hispanic	Multiracial	Caucasian	Other	Total
Scholarship Recipients	54.6%	7.0%	4.0%	32.3%	2.2%	100%
Public Applicant, Non-Recipients	64.1%	7.6%	6.1%	19.1%	3.1%	100%
Public Non-Applicants	74.0%	4.5%	3.8%	14.4%	3.2%	100%
Total	61.6%	6.3%	4.2%	25.3%	2.6%	100%

FIGURE 3. Minority Status Late Second Grade



4.1.4 Summary of Findings Related to Question One

From beginning first grade through late in their second-grade year, scholarship students, public school applicant/non-recipient students, and non-applicant public school students appear to come from families of roughly similar economic means. Significant differences among the groups on the estimate of income available for this study were found only between scholarship and non-applicant students at the end of the second grade. Similarly consistent were findings across the three data collection points indicating that scholarship students were more likely to be Caucasian than public school students whose parents did not apply for a scholarship.

Student mobility during first grade was similar across the three groups. Roughly 2.5% of students in each group changed schools at least once during that year. However, by the end of second grade, scholarship students had moved significantly less often than public school non-applicants. It is noteworthy that this is the case even though 46 scholarship students in the sample were forced to change schools between first and second grade when the two HOPE schools became community or charter schools[19]

4.1.5 Differential Entry of Scholarship Students

Embedded within the question of student characteristics are two relevant subquestions. First, are students who enter the Scholarship Program later (e.g., as second graders rather than first graders) different from those who enter earlier? To address this question, two sets of multivariate analyses of variance were conducted comparing the demographic characteristics and achievement of students who entered the program as: (i) kindergartners in autumn, 1997, (ii) first-graders in autumn, 1998, or (iii) second-graders in autumn, 1999. A second subquestion is, are students who leave the program after one or more years different from those who remain? Because data could be

19. Of the 46 students, 25 moved into other private schools, 8 moved into public schools, and the remainder could not be located in either Scholarship Office or Cleveland Public School records. It can be assumed that many, perhaps all of these 13 students continued to attend HOPE schools.

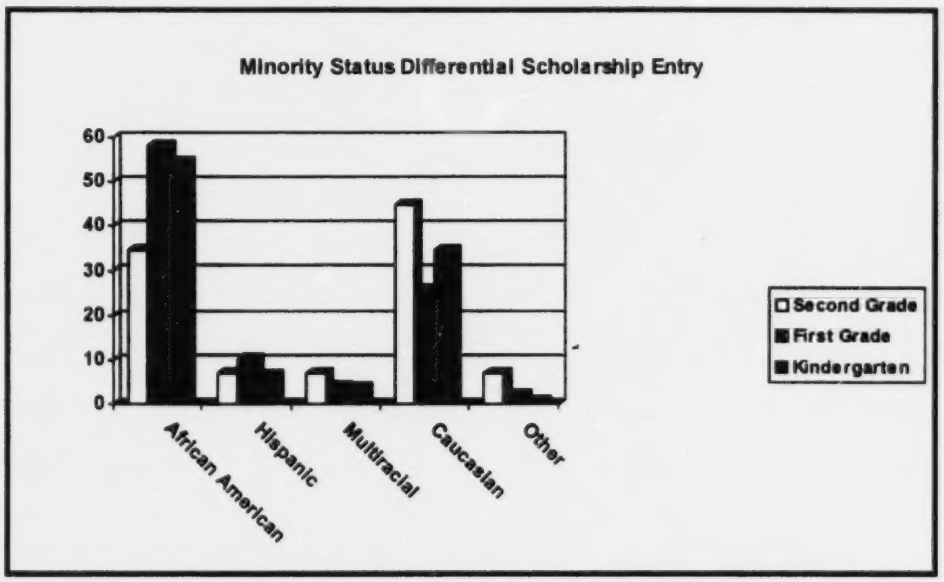
obtained on only 12 students who had left the program, it is not yet possible to answer this question. However, future evaluation activities are intended to provide sufficient data to do so.

Demographic data (gender, minority status, estimated meal code, and mobility) were available for 396 second grade scholarship students. Of these, 246 had begun the program as kindergartners, 121 as first graders, and 29 as second graders. On the basis of multivariate analysis (Pillai's Trace = .043, p = .030), separate univariate analyses were conducted for each variable. Descriptive statistics on these variables are presented in Table 9, and expanded minority status is depicted in Figure 4.

TABLE 9. Demographic Data for Scholarship Students of Differential Entry

Entry		Variable			
		Minority Status	Gender	Estimated Meal Code	Mobility
Second Grade	Mean	1.59	1.34	1.97	.069
	S.D.	.825	.480	.865	.260
First Grade	Mean	1.26	1.60	1.73	.110
	S.D.	.438	.490	.856	.340
Kindergarten	Mean	1.35	1.53	1.74	.120
	S.D.	.478	.500	.827	.320
Total	Mean	1.34	1.54	1.75	.11
	S.D.	.505	.500	.839	.320

FIGURE 4. Minority Status Differential Scholarship Entry



Analyses of variance indicated no significant differences among the groups for estimated meal code or mobility ($F [2, 393] = 1.12$ and .31, $p = .362$ and .735, respectively). However, significant differences were indicated for minority status and gender ($F [2, 393] = 5.27$ and 3.03, $p = .006$ and .049, respectively). Follow-up pairwise comparisons indicated that students who entered the program in second grade were significantly less likely to be of minority status than those who entered either as kindergartners or first graders. Similarly, second grade entrants were significantly more likely to be male than were students who entered the program as first graders.

Differential entry comparisons were also made on students' academic achievement each year. Multivariate analyses on the four measures of achievement were conducted for each of three testing episodes. The results of these analyses are described separately below.

Early first grade data were available for 32 students who would enter the Scholarship Program the next year (as second graders),[20] 134 students who had just entered the program, and 272 students who had entered the program as kindergartners the prior year. Descriptive statistics for the students on each of the four achievement measures obtained in fall, 1999 are reported in Table 10. Multivariate analysis indicated no significant differences (Pillai's Trace = .025, p = .196) in achievement among the groups. Thus, univariate analyses were not conducted.

TABLE 10. Fall, 1999 Achievement Data for Differential Entry

Entry		Measure			
		Reading	Language	Math	Total
Second Grade	Mean	550.75	558.31	513.25	540.66
	S.D.	5.718	6.622	6.050	5.017
First Grade	Mean	544.36	546.19	494.94	528.74
	S.D.	2.794	3.236	2.971	2.452
Kindergarten	Mean	546.54	549.94	502.07	532.480
	S.D.	1.961	2.271	2.086	1.721

Achievement data from late first grade were available for 406 scholarship students. This included 30 second grade students who would enter the program the following autumn, 123 first grade entrants, and 253 kindergarten entrants. Descriptive and multivariate analyses again were applied to these data. Descriptive statistics are presented in Table 11. No significant differences were indicated among the groups on achievement across the measures (Pillai's Trace = .025, p = .248).

20. Of the students who entered the Scholarship Program as second graders and for whom first grade achievement data are available, all had attended private school during first grade.

TABLE 11. Spring, 1999 Achievement Data for Differential Entry

Entry		Measure			
		Reading	Language	Math	Total
Second Grade	Mean	573.73	576.57	529.07	559.80
	S.D.	6.985	9.222	6.851	6.880
First Grade	Mean	562.50	558.72	513.48	546.55
	S.D.	3.450	4.554	3.383	3.398
Kindergarten	Mean	564.68	563.20	516.69	547.39
	S.D.	2.405	3.175	2.359	2.369

Second grade achievement data were available for 513 scholarship students (61 second grade entrants, 148 first grade entrants, 304 second grade entrants). Multivariate analysis of variance indicated no significant differences among the groups on the four achievement measures (Pillai's Trace = .014, p = .519). However, descriptive statistics are presented in Table 12.

TABLE 12. Spring, 2000 Achievement Data for Differential Entry

Entry		Measure			
		Reading	Language	Math	Total
Second Grade	Mean	602.12	596.62	557.66	585.56
	S.D.	4.53	5.11	4.42	4.10
First Grade	Mean	599.29	589.61	549.01	579.35
	S.D.	2.91	3.28	2.84	2.63
Kindergarten	Mean	597.65	594.31	551.46	581.26
	S.D.	2.028	2.288	1.979	1.838

It appears that scholarship students who enter the program at different times do not differ significantly on their entering level of achievement, prior mobility, or family income. However, students who entered the program most recently, as second graders, are significantly more likely to be male than students who entered in first grade,

and significantly more likely to be Caucasian than students who entered in either kindergarten or first grade.

4.2 Question Two

What are the characteristics of the classrooms and teachers with whom scholarship students work in private schools, and how do they compare with the characteristics of classrooms and teachers in public schools?

Data for Question 2 were analyzed for each testing episode, using classrooms as the unit of analysis rather than teachers. Comparisons were made between public and private schools at each of three times and are organized thusly below.

4.2.1 Autumn, 1998 (early first grade)

Table 13 presents descriptive statistics on classroom and teacher characteristics for the public and private schools attended by target students during autumn, 1998.

TABLE 13. Classroom and Teacher Characteristics, 1998

	Class Size		Highest Degree		Years Experience	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
Public N=78	23.30	.484	3.05	1.033	12.16	.095
Private N=48	24.63	.617	2.51	1.316	12.25	.121
Total	23.97	.392	2.78	.077	12.20	.836

Multivariate analysis of variance (MANOVA) suggested significant differences across the three classroom/teacher variables (Pillai's Trace F=5.049, p=.002). Subsequent univariate analyses of variance indicated no significant difference in class size between public and private schools (F [1, 124] = 2.859, p=.093), and no significant difference in teacher years of experience (F [1, 124] = .003, p=.956). Both public and private school classes included approximately 23 children during the autumn of first

grade, and the teachers in both sets of schools had just over 12 years of experience. However, public school teachers were significantly more likely to have completed graduate coursework than teachers in private schools ($F[1, 124] = 8.779$, $p=.001$). As can be seen in Table 14, substantially more public school teachers had completed upper level graduate degrees, particularly Master's Degrees, than private school teachers.

TABLE 14. Highest Teacher Education Level (in percent)

	Certified Only	B.A./B.S. Degree	B.A./B.S. Plus	M.A./M.S. Degree	M.A./M.S. Plus	Doctorate
Public School Teachers		35.9%	29.5%	30.8%	2.6%	1.3%
Private School Teachers		65.3%	18.4%	16.3%		
Total		47.2%	25.2%	25.2%	1.6%	.8%

Somewhat surprisingly, these teacher/classroom factors explained only a small portion of student academic achievement. Collectively, the three variables explained less than 10% of variance in academic achievement ($R^2 = .080$), although this was statistically significant.

4.2.2 Spring, 1999 (late first grade)

Table 15 presents descriptive statistics for teacher and classroom characteristics, aggregated by classroom, for the period spring, 1999.

TABLE 15. Classroom and Teacher Characteristics, 1999

	Class Size		Highest Degree		Years Experience	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
Public N=75	23.49	.484	3.02	.097	12.62	1.042
Private N=47	23.83	.612	2.44	.122	12.95	1.316
Total	23.66	.390	2.73	.078	12.79	.839

As during autumn of this same academic year, multivariate analysis of variance indicated significant differences in the variables across the public and private schools (Pillai's Trace F = 5.099, p=.002). Subsequent univariate analyses of variance found no significant difference in public and private school class sizes (F [1, 122] = .183, p = .670) and no significant difference in years of experience between public and private school teachers (F [1, 122] = .038, p = .846). And, as with the earlier autumn data, ANOVA revealed significant differences in highest degree completed between public and private school teachers (F [1, 122] = 13.696, p = .000). Again, and as shown in Table 16, public school teachers were much more likely to have completed at least Master's level coursework than private school teachers.

TABLE 16. Highest Teacher Education Level (in percent)

	Certified Only	B.A./B.S. Degree	B.A./B.S. Plus	M.A./M.S. Degree	M.A./M.S. Plus	Doctorate
Public School Teachers N = 75		42.3%	26.8%	31.0%		
Private School Teachers N = 47		70.2%	17.0%	12.8%		
Total		53.4%	22.9%	23.7%		

Data on these classroom/teacher variables indicated that they did not explain a significant portion of variance in student achievement, either individually or collectively (R^2 = .055, p = .085).

4.2.3 Spring, 2000 (late second grade)

Data on these classroom and teacher variables were again collected during spring of 2000, when students in the sample were nearing the end of second grade. Descriptive statistics on class size, teacher education level, and teacher experience are presented in Table 17.

TABLE 17. Classroom and Teacher Characteristics, 2000

	Class Size		Highest Degree		Years Experience	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
Public N=75	23.15	.515	3.11	.102	13.79	1.116
Private N=43	22.02	.681	2.44	.135	13.48	1.474
Total	22.58	.427	2.77	.085	13.63	.925

Multivariate analysis of variance indicated significant differences across the three variables. Follow-up univariate analyses of variance suggested no significant differences between public and private school teachers in terms of class size (F [1, 116] = 1.758, p = .187) or years of experience (F [1, 116] = .028, p = .868), but a significant difference favoring public school teachers in terms of education level (F [1, 116] = 15.427, p = .000). As Table 18 indicates, public school teachers were more likely to have completed graduate coursework, at least through the Master's Degree, than were private school teachers.

BEST COPY AVAILABLE

TABLE 18. Highest Teacher Education Level (in percentages)

	Certified Only	B.A./B.S. Degree	B.A./B.S. Plus	M.A./M.S. Degree	M.A./M.S. Plus	Doctorate
Public School Teachers N = 75		27.6%	46.1%	22.4%	3.9%	
Private School Teachers N = 43	16.3%	46.5%	18.6%	16.3%		2.3%
Total	5.9%	34.5%	36.1%	20.2%	2.5%	.8%

As in first grade, these teacher and classroom variables explain very little variance in students' academic achievement. Collectively, they explain only about 10% of total variance in achievement ($R^2 = .102$, $p = .005$), although this is statistically significant.

4.2.4 Summary of Findings for Question Two

Findings across the three testing episodes consistently indicated that differences in the characteristics of public and private school classrooms and teachers were minimal. Class size was not statistically different in either first or second grade. Students in both public and private schools were in classes of approximately 23 students. Teachers in both sets of schools were equally experienced, with between 12 and 13 years of experience. Teachers in both public and private schools were likely to have completed at least an undergraduate degree, although public school teachers were more likely to have completed graduate coursework than private school teachers.

4.3 Question Three

What is the impact of participation in the Scholarship Program on students' academic achievement?

In order to examine not only differences in achievement between and among the groups of students, but also to investigate patterns or trends in performance across time, mixed model analyses of variance were used in this portion of the project. To more fully understand how ongoing participation in the program might impact achievement, scholarship students were organized into two groups: students who entered the Scholarship Program as kindergartners and students who entered the program as first graders.[21] Separate analyses were conducted for each of the four academic measures provided by the Terra Nova (reading, language, mathematics, and total) across the three testing episodes. All analyses were conducted on scale scores to allow comparison across years as well as across student groups. Because none of the classroom, teacher, or student characteristics variables were found to explain a substantial portion of variance in student achievement, the present examination of achievement did not include these as covariates.

4.3.1 Reading Achievement

Descriptive statistics on students' adjusted reading achievement across the three years are presented in Table 19. Figure 5 presents these data graphically. Mixed model analysis of variance was used to analyze achievement data. The results of this analysis are presented in Table 20.[22]

21. Test data were also available for a group of 16 scholarship students who entered the program as second graders. However, because all of these students had attended a private school as first graders (i.e., before they entered the Scholarship Program), they are not included in the present analyses.
22. Estimates of sphericity were strong (Greenhouse-Geisser = 0.9530 – 0.9999). As a result, no adjustments to degrees of freedom were made.

TABLE 19. Student Achievement in Reading

		Autumn, 1998	Spring, 1999	Spring, 2000	Total
Two Year Scholarship	Mean	543.90	561.294	598.13	567.77
	S.E.	3.333	3.590	3.376	2.818
Three Year Scholarship	Mean	546.56	566.154	597.49	570.07
	S.E.	2.310	2.488	2.339	1.953
Applicant, Non-Recipient	Mean	541.80	571.142	599.69	570.88
	S.E.	2.480	2.670	2.511	2.096
Non-Applicant	Mean	534.77	563.286	596.51	564.86
	S.E.	1.871	2.015	1.895	1.582
Total	Mean	541.76	565.469	597.96	568.39
	S.E.	1.277	1.375	1.293	1.080

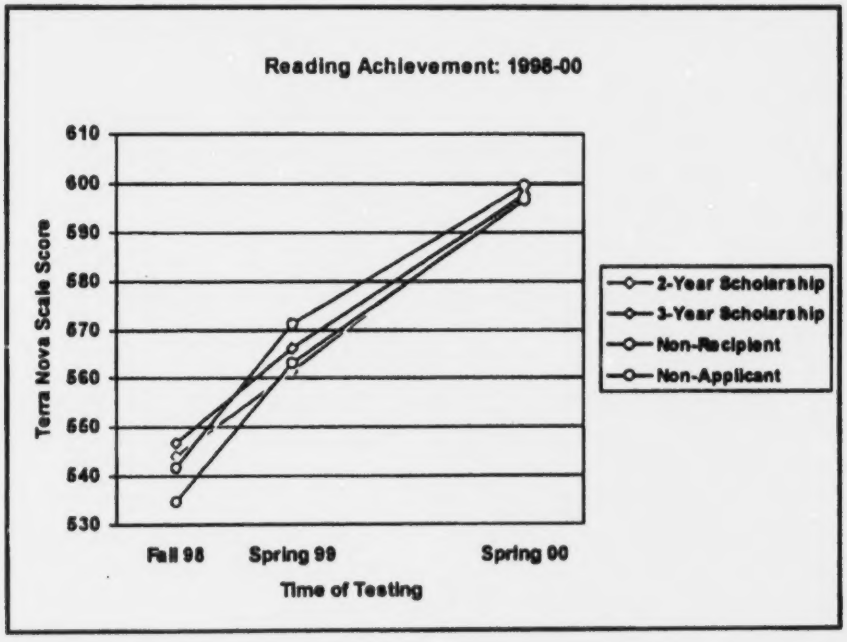
TABLE 20. Mixed Model Analysis of Variance on Students' Reading Achievement Scores: *Early First Grade through Late Second Grade*

Source					
	SS	df	MS	F	Sig
Within Subjects	2299843.39	1758			
Time	1182101.85	2	591050.93	937.56	.000
*Time * Status*	14520.44	6	2420.07	3.84	.001
Error	1103221.10	1750	630.41		
Between Subjects	2290946.29	878			
Scholarship Status	18070.11	3	6023.37	2.32	.074
Error	2272876.18				
Total	4590789.68	2626			

No significant main effects were found for group membership, suggesting that the four groups did not differ consistently from one another across the three testing episodes. However, significant effects were indicated both for time and for the interaction

effects of time and group membership. Each of these effects was further examined using pairwise comparisons.[23]

FIGURE 5. Reading Achievement: 1998-00



Follow-up analyses indicated that students in all groups improved significantly over each testing interval. Student reading scores were significantly higher at the end of first grade than at the beginning and were significantly higher at the end of second grade than they were at the end of first.

Of particular interest in the present evaluation are the interaction effects of scholarship status and time of testing. Again, follow-up pairwise comparisons were used to exam-

23. As noted earlier in this report, all pairwise comparisons were conducted using modified Tukey-Kramer techniques to control experimentwise error.

ine the nature of this interaction. The results of these comparisons indicated that non-applicant students were achieving at significantly lower levels in reading at the beginning of first grade. By the end of first grade, applicant/non-recipient students were achieving at significantly higher levels than 2-year scholarship students. And, by the end of second grade, there were no significant differences between or among any of the groups.

4.3.2 Language Achievement

Using students' scaled scores from the language portion of the Terra Nova, a second set of mixed model analyses was conducted. Descriptive statistics for the four groups are presented in Table 21 and graphically depicted in Figure 6. The results of analysis of variance are presented in Table 22.

TABLE 21. Student Achievement in Language: *Early First Grade through Late Second Grade*

		Autumn, 1998	Spring, 1999	Spring, 2000	Total
Two Year Scholarship	Mean	547.64	557.94	585.95	563.84
	S.E.	3.649	4.370	3.828	3.194
Three Year Scholarship	Mean	552.71	565.21	595.90	571.27
	S.E.	2.529	3.028	2.652	2.213
Applicant, Non-Recipient	Mean	542.19	566.76	589.15	566.04
	S.E.	2.707	3.242	2.840	2.370
Non-Applicant	Mean	539.82	562.99	582.63	561.81
	S.E.	2.045	2.449	2.145	1.790
Total	Mean	545.59	563.23	588.41	565.74
	S.E.	1.397	1.673	1.465	1.223

TABLE 22. Mixed Model Analysis of Variance on Students' Language Achievement Scores: *Early First Grade through Late Second Grade*

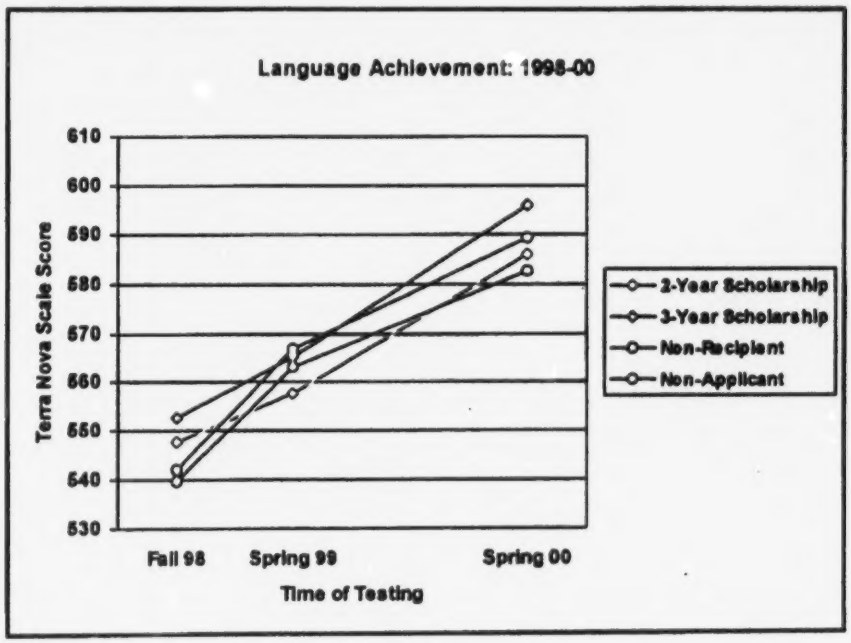
Source					
	SS	df	MS	F	Sig
Within Subjects	2280957.19	1762			
Time	688872.00	2	344436.00	383.85	.000
*Time * Status*	18172.87	6	3028.81	3.38	.003
Error	1573912.32	1754	897.33		
Between Subjects	2290946.29	878			
Scholarship Status	37888.05	3	12629.35	3.786	.010
Error	2925184.06	877	3335.44		
Total	5244029.3	2642			

In language, significant main effects were again found for time, with all students improving significantly across the testing episodes. Similarly, there were significant differences between and among the groups, despite time of testing. Pairwise comparisons revealed that 3-year scholarship students (i.e., those who began the program in kindergarten) had significantly higher language scores than public school non-applicants. However, there were no other significant group differences.

The interaction effects of time and group membership were also found to be significant. Follow-up pairwise comparisons found that 3-year scholarship students were achieving at significantly higher levels than both applicant/non-recipients and non-applicants at the beginning of first grade. There were no significant differences between 2-year scholarship students and any other group at this point. No significant differences were found between groups by the end of first grade. However, by the end of second grade, 3-year scholarship students were performing at significantly higher levels than non-applicant students and 2-year scholarship students (i.e., those who entered the program in first grade).

BEST COPY AVAILABLE

FIGURE 6. Language Achievement: 1998-00



4.3.3 Mathematics Achievement

Scaled scores from the mathematics portion of the Terra Nova were subjected to mixed model analysis of variance. Descriptive statistics by group and testing episode are presented in Table 23 and graphically depicted in Figure 7. The results of analysis of variance are presented in Table 24.

TABLE 23. Student Achievement in Mathematics: *Early First Grade through Late Second Grade*

		Autumn, 1998	Spring, 1999	Spring, 2000	Total
Two Year Scholarship	Mean	494.10	513.80	548.10	518.97
	S.E.	3.302	3.667	3.458	2.986
Three Year Scholarship	Mean	503.03	517.91	551.97	524.30
	S.E.	2.297	2.551	2.406	2.077
Applicant, Non-Recipient	Mean	492.50	518.56	555.41	522.15
	S.E.	2.410	2.676	2.524	2.179
Non-Applicant	Mean	490.84	516.06	551.42	519.44
	S.E.	1.804	2.004	1.889	1.632
Total	Mean	495.34	516.58	551.73	521.22
	S.E.	1.256	1.395	1.315	1.136

TABLE 24. Mixed Model Analysis of Variance on Students' Mathematics Achievement Scores: *Early First Grade through Late Second Grade*

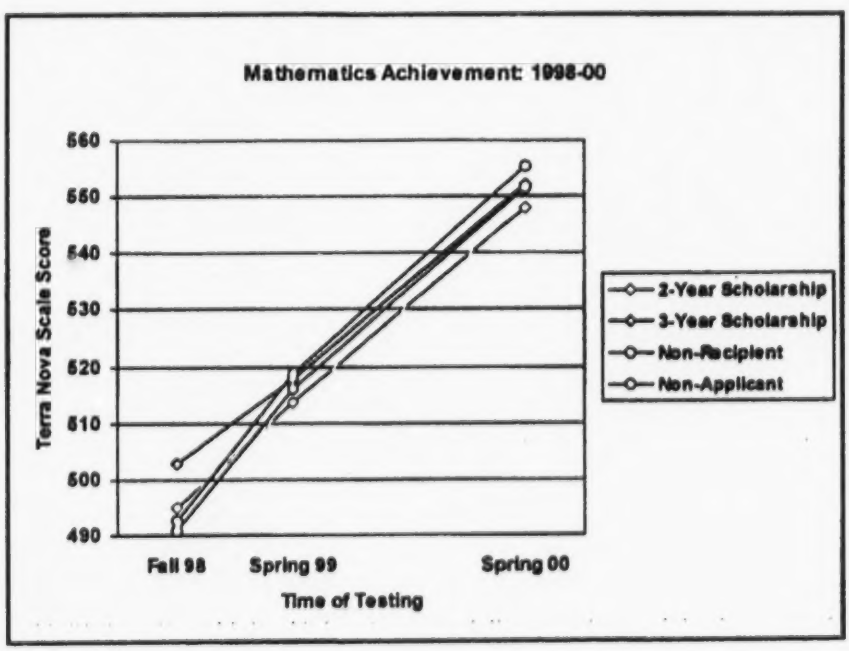
Source					
	SS	df	MS	F	Sig.
Within Subjects	2089981.25	1758			
Time	1188120.73	2	594040.36	1173.25	.000
*Time * Status*	15769.85	6	2628.308	5.191	.000
Error	886090.67	1750	506.34		
Between Subjects	4582877.12	878			
Scholarship Status	11754.24	3	3918.08	1.382	.247
Error	2492895.87	875	2835.59		
Total	2481141.63	2636			

Analysis of mathematics scale scores indicated a pattern of significant student improvement across the three testing episodes. Unlike the results for language scores, there are no significant between group differences that are independent of time of test-

ing. However, a significant interaction between time of testing and group membership is indicated.

Follow-up pairwise comparisons indicated that 3-year scholarship students were achieving at significantly higher levels than students in any other group at the beginning of first grade, and that the three remaining groups were not significantly different. The groups were achieving at equal levels by the end of first grade but were again different by the end of second grade. At the final testing point, applicant/non-recipients were achieving at significantly higher levels than 2-year scholarship students, with no other significant differences found among the groups.

FIGURE 7. Mathematics Achievement: 1998-00



4.3.4 Total Achievement

Scores on the Terra Nova are also provided to reflect total achievement across the content measures. These total scale scores were analyzed using mixed model analysis of variance and appropriate post hoc procedures. Descriptive statistics on these data are presented in Table 25 and graphically depicted in Figure 8. Results of the analysis of variance are presented in Table 26.

TABLE 25. Student Achievement Total Score: *Early First Grade through Late Second Grade*

		Autumn, 1998	Spring, 1999	Spring, 2000	Total
Two Year Scholarship	Mean	528.81	545.53	576.41	550.25
	S.E.	2.936	3.507	3.121	2.791
Three Year Scholarship	Mean	533.36	548.92	582.96	555.08
	S.E.	2.056	2.456	2.186	1.955
Applicant, Non-Recipient	Mean	526.70	553.12	582.58	554.13
	S.E.	2.215	2.645	2.355	2.106
Non-Applicant	Mean	521.66	547.81	577.11	548.86
	S.E.	1.668	1.993	1.774	1.586
Total	Mean	527.63	548.85	579.76	552.08
	S.E.	1.133	1.353	1.204	1.077

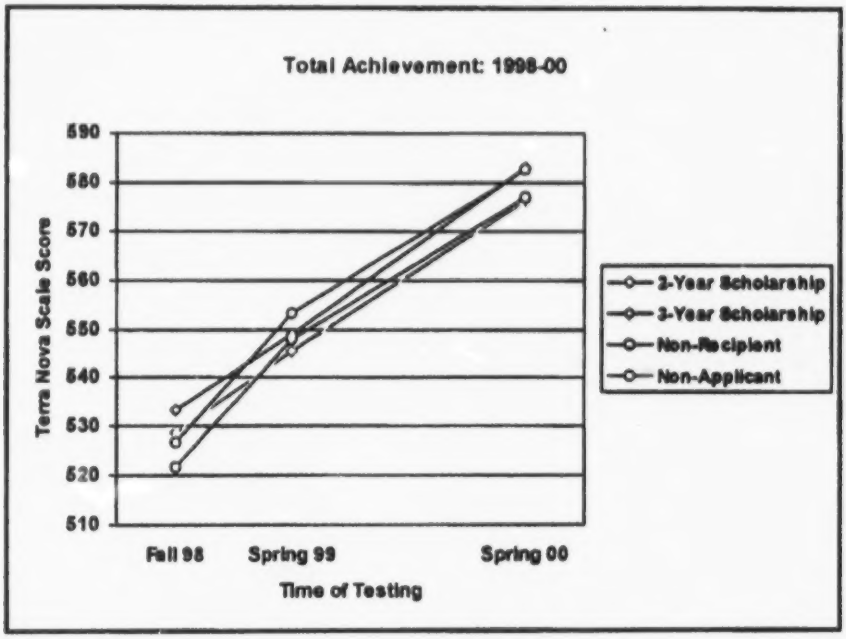
TABLE 26. Mixed Model Analysis of Variance on Students' Total Scores: *Early First Grade through Late Second Grade*

Source					
	SS	df	MS	F	Sig.
Within Subjects	1568671.26	1626			
Time	950803.79	2	475401.90	1267.221	.000
*Time * Status*	10869.68	6	1811.61	4.829	.000
Error	606997.79	1618	375.15		
Between Subjects	1966339.61	812			
Scholarship Status	18761.92	3	6253.97	2.598	.051
Error	1947577.69	809	2407.39		
Total	3535010.87	2438			

Analysis of total scale scores again indicates a significant effect for time of testing, with students in all groups improving significantly between each testing episode. No main effect for group status was found, but a significant interaction between time of testing and group membership was indicated.

Post hoc comparisons revealed that 3-year scholarship students were achieving at significantly higher levels on this measure than students in any other group at the beginning of first grade. Further, public school students who had not applied for a scholarship were achieving at significantly lower levels than any other group early in first grade. By the end of first grade, applicant/non-recipients were achieving at significantly higher levels than 2-year scholarship students, but all other groups were similar. By the end of second grade, 3-year scholarship students and applicant/non-recipients were achieving at significantly higher levels than both non-applicants and 2-year scholarship students.

FIGURE 8. Total Achievement: 1998-00



4.3.5 Summary of Findings Related to Question Three

The four, independent mixed model analyses reveal a pattern in the data across portions of the test. In general, students who entered the Scholarship Program as kindergartners began first grade achieving at higher levels than students in other groups. Relatedly, public school students whose families had not applied for a scholarship entered first grade achieving at lower levels than students in the other groups. By the end of first grade, the groups were mostly similar in level of achievement, although the achievement of applicant/non-recipients improved substantially more than that of other students. By the end of second grade, 3-year scholarship students again were performing at somewhat higher levels than other students and, at least in total score, the gains made by applicant/non-recipient students are maintained.

Indiana Center for Evaluation

5 CONCLUSIONS

The findings presented in the preceding section have been drawn from data collected on students who were in the Scholarship Program or attending public schools as first graders in the autumn of 1998. These students include private school students who are scholarship recipients, public school students who applied for but did not receive a scholarship, and public school students whose families did not apply for a scholarship. In addition, to these students for whom sufficient data are available, a small group of public school students can be identified who were offered and perhaps used a scholarship, but who have chosen to return to public schools. However, this group remains too small to be included in current analyses.

In addressing three primary evaluation questions, the present findings represent the first of what are anticipated to be a series of results drawn from ongoing evaluation of the Cleveland Scholarship Program. They are based upon achievement, demographic, and classroom data drawn over a two-year period, from autumn, 1998 through spring, 2000. This period reflects students' first and second-grade years in their public or private schools. Thus, they are only early findings associated with what will be students' extended school experiences. And, they must be considered directly representative only of students, teachers, and schools associated with the Cleveland Scholarship Program during this period. While it is tempting to generalize the current findings, either by assuming they reflect future results or by using them to predict results from choice programs in other settings, it is completely inappropriate to do so.

It is with these caveats that we begin a discussion of what we believe are reasonable conclusions to be drawn from the current findings. The remainder of this section of the report attempts to put findings associated with each evaluation question into context, and concludes with a short summary across the questions.

5.1 Question One

What are the characteristics of students who participate in the Scholarship Program, and how do they compare with those who do not participate?

Over the two-year period, all groups in our sample consist nearly equally of males and females. Further, in first grade, scholarship and applicant/non-recipient students are nearly identical in terms of family income, gender, minority status, and mobility. These students' estimated meal code was at approximately the mean for all students (roughly $3,242 per person), they were between 22% and 33% non-minority, and they were roughly equally male and female. Further, applicant/non-recipients were very similar to non-applicants in these ways. However, scholarship students were significantly less likely to be of minority status than were non-applicants, and additional differences between groups appear to develop by the end of second grade.

By late second grade, scholarship students continued to be less likely to be of minority status than non-applicants (67% versus 85%), were of significantly higher income (approximately $150 per person annually), and had changed school significantly less often. They were also found at this point to be significantly less likely to be of minority status than applicant/non-recipients (78% minority).

Thus, when viewed as a single group, scholarship students entered first grade different from the public school comparison students only in the extent to which they were of minority status. But the group characteristics change from beginning first grade to late second grade and result in several important differences, particularly between scholar-

ship students and non-applicants. Estimated income for non-applicants actually diminishes by the end of second grade while increasing slightly for scholarship students. Further, only 9.2% of scholarship students had changed schools during first or second grade, 15% of non-applicants had changed schools at least once during this period.

These seeming changes over time may reflect actual differences in the characteristics of the group. It is logical and reasonable that students whose parents were able to use a scholarship to choose the private school they would attend, would be less likely to change schools than families who did not choose. However, particularly for estimated meal code, used as a proxy for family income, it is likely that impreciseness of measurement explains some of the apparent change over time. It is also reasonable that some of these changes in group characteristics are the result of students entering or leaving groups within the sample. This seems particularly important in terms of students who entered the Scholarship Program over time.

Comparisons of the characteristics of students who entered the program at different times appear to support the contention that students constituting the scholarship group, at least, change in character over time. It was not possible to compare the characteristics of students who remained in the program with those who chose to leave. However, analyses indicate that students who entered the program at the beginning of second grade differed in important ways from those who entered the program earlier. The scholarship students are of roughly equal levels of achievement throughout the period. And, students who enter in kindergarten and first grade are not statistically different on any of the demographic variables. However, students who enter the program in second grade were significantly less likely to be of minority status (41%) than students admitted in either kindergarten or first grade (65% and 74%, respectively). Further, they were more likely to be male (66%) than first grade entrants (40%).

What remains to be seen is the pattern these group characteristics may take in coming years. If these data reflect a trend that continues for several years, the nature of students in the Scholarship Program may change and become less similar to students in public schools. It is additionally important to understand whether these changes are a result of differential student characteristics among those who enter the program in later grades, whether these reflect group changes due to attrition from the program, or some combination of these factors. In any event, the findings suggest that the scholarship students are not like their public school counterparts in all ways.

5.2 Question Two

What are the characteristics of the classrooms and teachers with whom scholarship students work in private schools, and how do they compare with the characteristics of classrooms and teachers in public schools?

In the earlier phase of our evaluation, we found that public and private school classrooms and teachers differed significantly in class size, teacher experience, and teacher education level. Over the first two years of the second phase, however, only teachers' education level is found to differ between the groups.

Both public and private school classrooms serve approximately 23 students in both first and second grades. Our earlier data indicated significantly, though slightly, larger classrooms in public schools (23.6 students public, 20.6 students private). Both public and private school teachers in the present sample were found to have over 12 years of teaching experience, whereas our earlier data revealed public school teachers to have significantly more experience than private school teachers (14.19 years public, 8.55 years private). However, one finding remained consistent: public school teachers were significantly more likely to have completed graduate coursework than were private school teachers. Between 64% and 72% of public school teachers had taken at least some graduate coursework, while only 30% to 54% of private school teachers had done so.

Thus, in the current sample, the classroom and teacher characteristics experienced by students in public and private schools are somewhat more similar than we had found before. The substantial financial incentives provided to public school teachers for completing graduate coursework likely explain this consistent difference with private school teachers. It is less clear why other differences, even those that were found in our earlier work, were not found in the present data. Public school teachers in our sample had nearly the same level of experience as those we studied in previous years. In contrast, the experience of private school teachers in the present sample was nearly four years greater than we had found before. Sample error might explain some of this apparent difference, though data on nearly all private and public school teachers were obtained each year. It is also possible that the difference in grade levels at which we collected data be a factor. In the earlier work, we collected data in third, fourth, and fifth-grade classrooms. In the current work, data were drawn from first and second grades. Perhaps, particularly in the private schools, teachers with greater experience tend toward the early grades, while less experienced teachers are assigned to upper elementary grades.

A final point is, we believe, noteworthy. The three teacher/classroom variables that were examined in this phase of the study were not substantially related to student achievement. In fact, these variables accounted for less than 10% of variance in student achievement gains over the two-year period. A lack of variance between the groups on these variables may explain some of this. Still, these factors have long been associated with varying student achievement levels, with an assumption that greater teacher experience, greater teacher education, and smaller class size promotes greater student learning. Class size reduction, in particular, rivals school choice as the most hotly and frequently debated educational issue. Unfortunately, our current data do not indicate that these factors influenced substantially the achievement of students in our sample.

5.3 Question Three

What is the impact of participation in the Scholarship Program on students' academic achievement?

Unlike our earlier work, which could include only a comparison group of students whose families had never applied for a scholarship, the present phase of the evaluation has included multiple comparison groups. In addition, all achievement data used in the present work was obtained through evaluation controlled and monitored administration of a consistent achievement test series. As a result, the consistency and comparability of the data, both across students and across years, are likely to be enhanced.

The analytic approach that we took allowed us to examine three aspects of influence on students' achievement in each of the four areas measured by the test. In each analysis, we are provided with information about the effect of time on students' performance, independent of which group (i.e., scholarship, applicant/non-recipient, or non-applicant) they are in. In other words, we answer the question, does student performance change significantly over time, independent of group membership? Because we would hope that students' learning in schools is cumulative, we can assume that their test scores would increase as they experience more formal schooling. And, our data bear this out. Students' academic performance in each of the four areas improved significantly from beginning first grade to the end of first grade and from the end of first grade to the end of second grade. This was true for all students, independent of their scholarship status and suggests that students benefited significantly from their schooling, whether in private or public schools and whether they were scholarship, applicant/non-recipient, or non-applicant students. This pattern of improvement can easily be seen in graphs presented in Figures 5, 6, 7, and 8. The achievement score of each group increased at each testing episode.

The analytic techniques we used also provided information about the effects of group membership on students' achievement, independent of time. The question addressed

here is, are there consistent differences at each time of testing between or among the groups based on scholarship status? With the exception of language, in which scholarship students who began the program in kindergarten had consistently higher scores than did non-applicants, no pattern of achievement difference was found based on scholarship status. Again, this largely disordinal pattern is observable in the graphs presented earlier.

For our purposes, the most important information provided by the analytic technique is an indication of the interaction of scholarship status and time. If the Scholarship Program influences student achievement positively, a significant interaction of time and group would be seen as students in the scholarship group begin to improve at greater levels than students in other groups. In each of the four achievement analyses, an interaction effect was found to be significant. However, the trend of this interaction was not consistent across the four measures nor did it favor any one group. In general, students who entered the Scholarship Program as kindergartners (and thus had participated for three years by the end of second grade) began first grade achieving at higher levels than other students. And, public school students whose parents had never applied for a scholarship were often achieving at lower levels than other students when they began first grade.

By the end of first grade, earlier between group differences were reduced or eliminated. Non-applicants improved comparatively in each of the four areas during their first-grade year, and public school applicant/non-recipients improved similarly, though not as consistently or dramatically. Graphical presentation of students' achievement in each of the four areas indicates that the pattern of general improvement between students attending public schools (applicant/non-recipients and non-applicants) appears to differ from that of scholarship students. By the end of second grade, some differences again begin to emerge. By this point, 3-year scholarship stu-

dents and applicant/non-recipients had improved more dramatically than either 2-year scholarship students (those who entered the program in first grade) or non-applicants.

The patterns of achievement gain over time and between groups are not clear, and data are only available for the first two years of students' formal schooling. As a result, it is much too early to draw definitive conclusions about the impact of the voucher program on students' achievement. As we have maintained throughout our work, and as research on vouchers in Milwaukee and other cities has indicated, the effects of schooling, and thus of the Scholarship Program, are cumulative and incremental. Further, research of other voucher programs indicates that the positive effects of choice are more likely as students move beyond the primary or early elementary grades.[24] However, some basic, though tentative, conclusions may be suggested.

First, students who enter the Scholarship Program as kindergartners seem both to have entered first grade with an academic advantage and to be more positively impacted by the program than those who enter as first graders. It is not clear whether this represents the impact of the Scholarship Program or differences in the characteristics of students who constituted each group in our sample. For example, the scholarship may have made enrollment in kindergarten easier or more likely and, thus, the 3-year scholarship students in our sample may have had more experience in formal schools than students in the other groups. Or, parents who were sufficiently interested and motivated to obtain a scholarship for their children to attend private kindergarten may be more involved, interested, and supportive of their children's education than other parents. Unfortunately, resources for this period of the evaluation did not allow collection of data that might resolve this issue. However, such data are intended for collection in coming years.

24. See Witte (2000); Howell et al. (2001).

A second general conclusion associated with students' achievement is that students, no matter whether they participate in the Scholarship Program or whether they attend public or private schools, improved significantly over their first two years of school. While this may seem only reasonable, recent criticisms of schools (public schools by voucher advocates, private schools by voucher opponents) often give the impression that students do not gain academically from their school experiences. While the extent or efficiency of this gain may be debated, our current data indicate that these first and second-grade students did, in fact, benefit significantly from schooling.

Third, and perhaps most speculative, is the conclusion that differential patterns of improvement between public and private school students seem to be reflected in the data. Public school students generally made greater academic improvement during first grade than did scholarship students. Public school improvement from beginning to end of first grade was nearly 30 scale score points, whereas private school students' improvement during this period was approximately 15 points. From the end of first grade to the end of second, students' mean achievement increase across the four measures was nearly equal. The improvement of public school students during first grade, particularly as compared with the scholarship students during this period, may be due to regression toward the mean resulting from their relatively low entering levels of achievement. It may also reflect an emphasis in public schools on helping these young children "make up" for real or perceived deficiencies as they begin school. Whatever the reason for this difference during first grade, the nearly equal levels of improvement during second grade suggest that something critical differed in the experiences of the scholarship and public school students during first grade. This difference may have been institutional, programmatic, environmental, or random, and future data collection should address this.

5.4 Summary

Across the evaluation findings from these early years of students' experience in the Scholarship Program, there is little that is provocative. Scholarship students enter the program with somewhat different academic and demographic characteristics than students who attend public schools. Public school teachers are more likely to have completed graduate coursework, particularly up to the master's level, than are private school teachers. Class size and teacher experience, at least for the samples of classrooms we examined, were quite similar between public and private schools. And student academic achievement, a factor that is watched closely by those on both sides of the voucher issue, presented no clear or consistent pattern that can be attributable to program participation.

The findings that we have presented, and the conclusions that we have attempted to draw from it, are extremely tentative. As was the intention of the independent technical panel that advises the project, they represent only the first of what are to be multiple findings drawn on students over a period of several years. As a result, the findings are incomplete and understandably, inconclusive. Yet, they are informative about the ways in which a choice program, or schools generally, influence the early school experience of children. Advocates and opponents of the Cleveland Scholarship Program will each find information from the present study useful and supportive of their position. However, the findings also suggest that what is often assumed ideologically to be true for all students, or all schools and teachers, or all voucher programs, may not necessarily be true at all. Only the continued generation of empirical evidence, collected in the most rigorous and neutral manner possible, can help policy makers, educators, and parents make informed and effective educational choices for their children. It is in this respect that the publicly-funded pilot program in Cleveland and the ongoing evaluation that it includes are critical to the continued discussion surrounding the voucher issue.

Indiana Center for Evaluation

6 FUTURE EVALUATION ACTIVITIES

The present project is longitudinal in nature, and it is expected to continue to collect data on students and their classrooms for several years. As noted early in this report, the unique status of the Cleveland program and paucity of research on publicly-funded voucher programs increases the importance of the present, ongoing work. The exact nature of the future evaluation activities will be reflexive to new information needs and resource availability. However, we anticipate that future activities will include the following:

- Achievement data will continue to be collected annually on students in the cohort. The Terra Nova will be administered to students in spring of each future year. In addition to minimizing attrition from the existing sample of students, particular attention will be focused on attempting to obtain data on the small number of students who choose to leave the program over time. The next administration of the achievement tests is late April, 2002 for fourth grade students.
- Classroom and teacher data will be collected as in the past, but supplemented to include information about: the amount of time spent on core subjects and resources (e.g., textbooks, workbooks, manipulatives, etc.) available for teaching these subjects. These additional data are scheduled to be collected for the first time during teacher interviews conducted by the test proctors in late April, 2002.
- Arrangements are currently being made to access students' attendance data through Cleveland Public School and Cleveland Scholarship Office records. These data are reported in substantially different ways across the schools, but will be analyzed in an attempt to identify reasonable and valid methods of making cross-school comparisons. This activity will not begin until the 2001-02 academic year.

- During the next two-year evaluation period, telephone interviews of randomly sampled parents of public and scholarship students will be conducted. These interviews will parallel the interviews conducted in spring, 1999, but will also examine issues that have arisen since that time (e.g., the impact of legal uncertainty of the program, reasons for pursuing or not pursuing a scholarship, perceptions of the type and quality of instruction offered their children, etc.).
- Also during the next two-year period, written surveys of teachers, principals, and students will be conducted to gain information on perceptions and awareness of the Scholarship Program. These surveys will be supplemented with a small number of focus groups.
- If resources allow, the evaluation will include examination of the financial, procedural, operational, and political impacts of the Scholarship Program on the public and private schools in Cleveland. Existing documents, personal interviews, and case studies will be used to attempt to identify and understand how the Scholarship Program may have influenced the educational context in Cleveland.

6.1 Closing Comments

Before closing, a point should be made about the context in which the Cleveland Scholarship Program has operated during the period covered by this report. While we have focused on our own data collection and findings associated with the relatively narrow evaluation questions, parents and educators in Cleveland have been making decisions about a Scholarship Program whose future has been extremely uncertain. Throughout the period of this project, the constitutionality of the program has been subject to legal challenge. During the summer of 1999, only days before the students in our sample were about to begin their second-grade year, a federal judge ruled the program unconstitutional and ordered that the Scholarship Program was immediately to be curtailed. Shortly thereafter this order was relaxed to allow scholarship students already admitted to the program to continue. Since then, a federal court panel has similarly ruled the program unconstitutional, but the program has been allowed to continue to operate as the case is appealed to the U.S. Supreme Court.

Though the evidence is only anecdotal, gathered from our informal conversations with parents and with public and private school educators, this uncertainty seems certain to have impacted the program, the schools, and the families that participate. Private schools, many of which are forced to turn away new students for lack of space or instructional personnel, are reluctant to commit to new capital investments or personnel until they feel the program will continue. Parents have been similarly reluctant to place their children into a program and new private school knowing that they would be forced to move their child if the program ends. It is in this atmosphere that the Cleveland Scholarship Program has operated, and it is this context of uncertainty that the data we have presented were collected. We cannot know to what extent uncertainty affected the families or children, and it is our hope that future data collection will allow us to do so. But it does remind us that the debate is not just about policy or politics – it is about children and families who must do the best they can in spite of the uncertainty.

THIS DOCUMENT CONCERNS THE CLEVELAND SCHOLARSHIP AND TUTORING PROGRAM AS ESTABLISHED PURSUANT TO OHIO REVISED CODE SECTION 3313.975, AND SHOULD NOT BE CONFUSED WITH THE CLEVELAND SCHOLARSHIP PROGRAM, INC.

Evaluation of the Cleveland Scholarship Program

1998 - 2000
Summary Report

Kim K. Metcalf, Ph.D., Director

September, 2001

Indiana Center for Evaluation

Smith Research, Suite 100
2805 East 10th Street
Bloomington, Indiana 47408-2601
http://www.indiana.edu/~iuice/

Kim K. Metcalf, Ph.D.
Director

812-855-4438
800-511-6575
Fax: 812-856-5890
iuice@indiana.edu

Indiana Center for Evaluation

THIS DOCUMENT CONCERNS THE CLEVELAND SCHOLARSHIP AND TUTORING PROGRAM AS ESTABLISHED PURSUANT TO OHIO REVISED CODE 3313.975, AND SHOULD NOT BE CONFUSED WITH THE CLEVELAND SCHOLARSHIP PROGRAM, INC.

EVALUATION OF THE CLEVELAND SCHOLARSHIP PROGRAM

1998-2000

EXECUTIVE SUMMARY

September, 2001

The current evaluation focuses on a cohort of children who entered school as kindergartners in 1998. The cohort consists of three groups of students: those who were awarded scholarships to attend private school, those who applied for a scholarship but who did not win an award (and subsequently stayed in public school), and finally, public school students who did not apply to the scholarship program. The objective of the evaluation is to follow this group of nearly 2,600 children as they make their way through elementary school.

The present report covers the period from autumn, 1998, when students were early in their first grade year, through late April 2000, when students were nearing the end of their second grade year. Drawing on achievement test data, classroom interviews, and records from public and private schools, the focus of this period has been on three questions:

1. What are the characteristics of students who participate in the Scholarship Program, and how do they compare with students who do not participate?
2. What are the characteristics of teachers and classrooms experienced by scholarship students, and how do they compare with those experienced by public school students?
3. What are the impacts of the scholarship program on students' academic achievement?

Based on the results of a variety of descriptive and inferential techniques, the preliminary findings suggest the following:

STUDENT CHARACTERISTICS

- Students who were awarded scholarships are more likely to be white (29.5%) than students who applied and did not win (21.3%) and students who did not apply (16.4%).

- Per capita income ($3,242) is not significantly different among families of scholarship students, students who applied and did not win, and students who did not apply.
- Scholarship students are less likely to have changed schools than their public-school counterparts by the end of second grade.

TEACHER AND CLASSROOM CHARACTERISTICS

- Teachers in scholarship and public schools are similar in terms of experience (mean = 12.2 years).
- Teachers in scholarship schools are less likely to have completed graduate degrees (16.3%) than their public-school counterparts (34.3%).
- Classroom size is comparable across scholarship and public schools (mean = 23.97 students).

STUDENT PERFORMANCE

- Students who entered the Scholarship Program as kindergartners were achieving at significantly higher levels than other students when they entered first grade.
- Students across all three groups (scholarship, applicant non-winners, and non-applicants) made substantial and statistically significant gains in performance between kindergarten and grade two.
- Gains made by public-school students in grade one generally closed the performance gap, but students enrolled in the Scholarship Program for the full three years continue to perform slightly, but statistically significantly, higher.

CONCLUSIONS

The current findings will likely be used to support arguments on both sides of the voucher issue, but the results are far too preliminary to be conclusive, and much remains to be learned. From an educational perspective, it may be useful and informative to examine the educational experiences that led to the substantial gains made by the public school students during first-grade. More generally, will trends develop in students' achievement as they move further through their schooling? Next steps in the evaluation will include further monitoring of student performance and additional, in-depth review of teacher and parent perspectives.

Indiana Center for Evaluation

1998-2000 SUMMARY REPORT

Evaluation of the Cleveland Scholarship Program was initiated by the Indiana Center for Evaluation in the spring of 1997, at the conclusion of the first operating year of the program. From the beginning, the evaluation has examined the impact of the Scholarship Program on the students, families, and schools that participate, and the impact on the public schools from which the scholarship students are drawn. In November 1998, the evaluation began following a cohort of first grade scholarship and public school children, with the intention of collecting longitudinal data on the impact of the program.

The present report details the findings of data drawn from the cohort from the time the students began first grade through the end of second grade. The sample under study includes three primary groups of students for whom sufficient data were available: (1) first grade *scholarship students* who began the program as kindergartners during 1996-97 or first graders during 1997-98 (N = 780); (2) first grade public school students whose families had applied for a scholarship during 1996-97 or 1997-98, but who were not selected to receive one through the random lottery or who received a scholarship, but chose not to use it (N = 612), which are referred to as *applicant/non-recipients*;[1] and (3) first grade public school students whose families had never applied for a scholarship (N = 1,233), referred to as *non-applicants*. When appropriate, students who accepted and began using a scholarship as second graders during 1999-00 were

1. During the period of this evaluation, records did not allow distinctions to be made between families who were awarded a scholarship but chose not to use and those that were not selected in the random lottery.

also included in analyses. As a result, the current project includes multiple comparison groups by which scholarship students' progress is weighed.

Three evaluation questions guided the project. These questions are:

1. What are the characteristics of students who participate in the Scholarship Program and how do they compare with students who do not participate?
2. What are the characteristics of the classrooms and teachers experienced by students who participate in the program, and how do they compare with those experienced by public school students?
3. What is the impact of participation in the program on students' academic achievement, and other relevant school related variables (e.g., attendance, behavior)?

During 1998-2000, the evaluation emphasized the identification of complete and representative groups of students who constitute the longitudinal sample and on collecting initial data on questions 1, 2, and 3 above. A detailed technical report has been prepared in which the statistical and methodological techniques of the project are explained and justified. The present report attempts to summarize the basic evaluation activities in ways that avoid technical language and that focus on the findings and their interpretation. Individuals who would like more information on the details of the study, including the specific measures used in collecting data and their operational definitions, are encouraged to obtain the longer technical report.

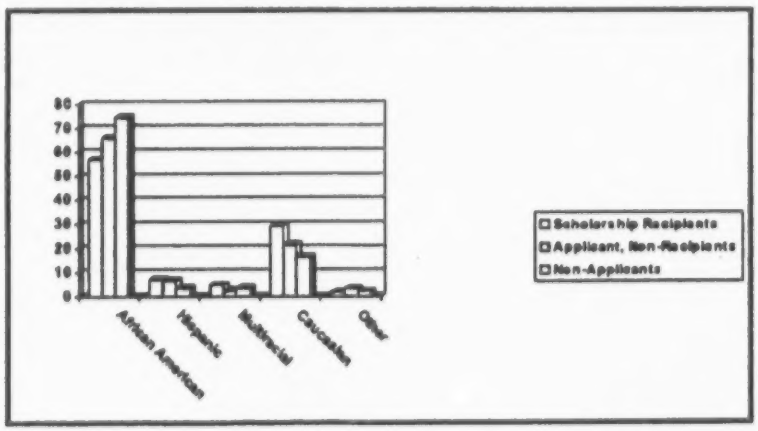
The remainder of this summary is organized around the three evaluation questions for which data were collected and analyzed. Findings are frequently presented graphically rather than in tables and supplemented with narrative explanations for clarity.

EVALUATION QUESTION ONE

What are the characteristics of students who participate in the Scholarship Program, and how do they compare with students who do not participate?

At the beginning of first grade, students in each of the groups were of similar gender and of similar family income, but differed in minority status.[2] Students in each of the groups were about equally male and female, and estimated per capita income was roughly $3,242. However, students who participated in the Scholarship Program were significantly less likely to be of minority status than were non-applicants. Only about 70% of scholarship students were of minority status in contrast to 83% of non-applicants. Applicant/non-recipients and non-applicants, however, were of roughly equal minority status (79% minority for applicant/non-recipients). Figure 1 graphically depicts minority status across groups.

FIGURE 1. **Minority Status Early First Grade**



2. Because no comparable measure of family income was available on all students, a proxy for this variable was computed using data from Cleveland Scholarship Office and Cleveland Public Schools records. This new variable, called Estimated Meal Code, was used in all analyses. Details of this variable and its calculation are explained in the detailed technical report.

Additional comparisons suggest that students who enter the Scholarship Program as second graders differed in important ways from those who entered in kindergarten or first grade. All scholarship students were of roughly equal levels of achievement throughout the period, and students who entered in kindergarten and first grade were not statistically different on any demographic variables. However, students who entered the program in second grade were significantly less likely to be of minority status (41%) than students admitted in either kindergarten (65%) or first grade (74%). Further, second grade entrants were more likely to be male than first grade entrants (40%).

Thus, students in the Scholarship Program are somewhat less likely to be of minority status than their public school peers, and this pattern continues and is more distinct as students enter the program as second graders. What remains to be seen is the pattern these group characteristics may exhibit in coming years. If these data reflect a trend that continues for several years, the nature of students in the scholarship program may change substantially and become less similar to students in public schools. It is additionally important to understand whether these changes are a result of differential student characteristics among those who enter the program in later grades, whether these reflect group change due to attrition from the program, or some combination of these factors. In any event, the findings suggest that the scholarship students are not like their public school counterparts in all ways.

EVALUATION QUESTION TWO

What are the characteristics of the classrooms and teachers experienced by students who participate in the Scholarship Program, and how do they compare with those experienced by public school students?

Public and private school classrooms experienced by the students in the sample serve approximately 23 students in first grade and nearly the same number in second grade.

Both public and private school teachers in the present sample were found to have over 12 years of teaching experience. However, public school teachers were significantly more likely to have completed at least some graduate coursework than were private school teachers. Between 64% and 72% of public school teachers had taken at least some graduate coursework, while only 30% to 54% of private schools teachers had done so.

Thus, in the current sample, the classroom and teacher characteristics experienced by students in public and private schools were more similar than we had found in our earlier work.[3] The substantial financial incentives provided to public school teachers for completing graduate coursework likely explain some of this consistent contrast with private school teachers. It is less clear why other differences, even those that were found in our earlier work, did not appear in the present data. Public school teachers in our sample had nearly the same level of experience as those we studied in previous years. In contrast, the level of experience of private school teachers in the present sample was nearly four years higher than we had found before. Sample error might explain some of this, though data on all private and public school teachers were obtained each year. It is also possible that the difference in grade levels at which we collected data are a factor. Perhaps, particularly in the private schools, teachers with greater experience tend toward the early grades, while less experienced teachers are assigned to the upper elementary grades.

A final point is noteworthy. The three teacher/classroom variables that were examined in this phase of the study were not substantially related to student achievement. In fact, these variables accounted for less than 10% of variance in student achievement gains over the two-year period. A lack of variance between the groups on these variables may explain the limited relationship with student achievement. Still, these fac-

3. K. Metcalf, W. Boone, P. Muller, P. Tait, and N. Stacey, *Evaluation of the Cleveland Scholarship Program: 1997-1999* (Bloomington, IN: Indiana Center for Evaluation, 1999).

tors have long been associated with varying student achievement levels, with an assumption that greater experience, greater teacher education, and smaller class size promotes greater student learning. Class size reduction, in particular, rivals school choice as the most hotly and frequently debated educational issue. Unfortunately, our current data do not indicate that these factors substantially influenced the achievement of students in our sample.

EVALUATION QUESTION THREE

What is the impact of participation in the Scholarship Program on students' academic achievement?

Graphical representations of students' performance in reading, language, mathematics, and total score are presented respectively in Figures 2, 3, 4, and 5.[4] The analytic approach allowed examination of three aspects of influence on students' achievement in each of the four areas measured by the test and can be explained in terms of these graphs. For each subject area or score, a result is provided that indicates the effect of *time* on student scores (i.e., do all of the lines change in roughly parallel ways across the three testing episodes?); the effect of *group* on student scores (i.e., does one or more of the lines differ from the others consistently across the three testing episodes?); and the interaction effect *of time and group* (i.e., does the performance of one or more groups change more or less than the others over time?).

4. Scholarship students were analyzed as two distinct groups based upon whether they entered the program in kindergarten or in first grade. As a result, there are four groups presented in the graphs.

FIGURE 2. Reading Achievement: 1998-00

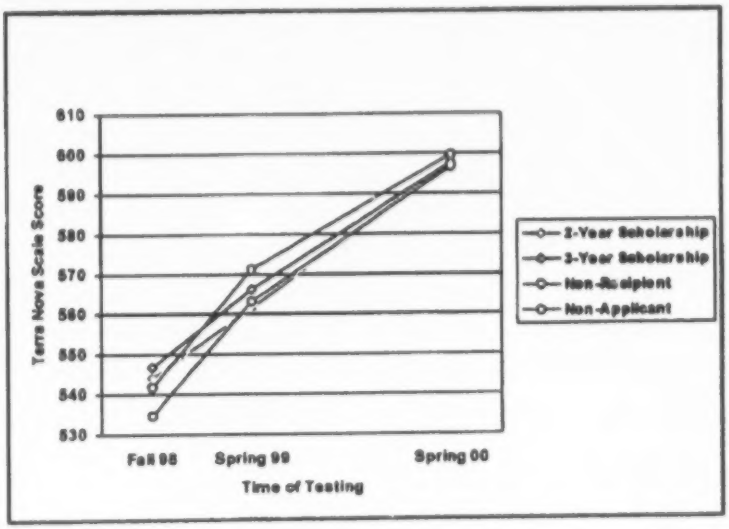


FIGURE 3. Language Achievement: 1998-00

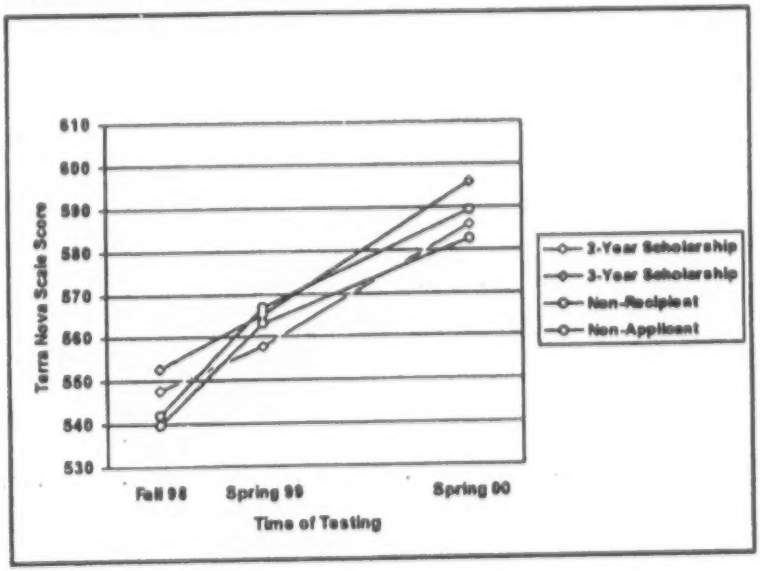


FIGURE 4. Mathematics Achievement: 1998-00

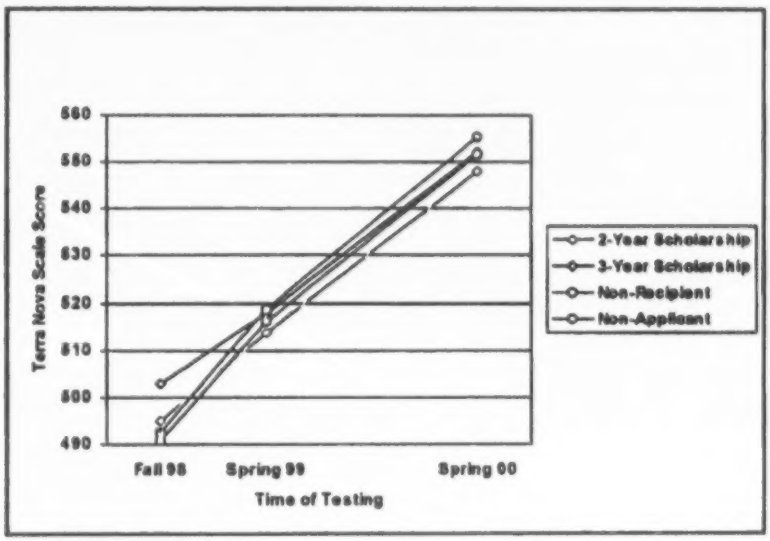
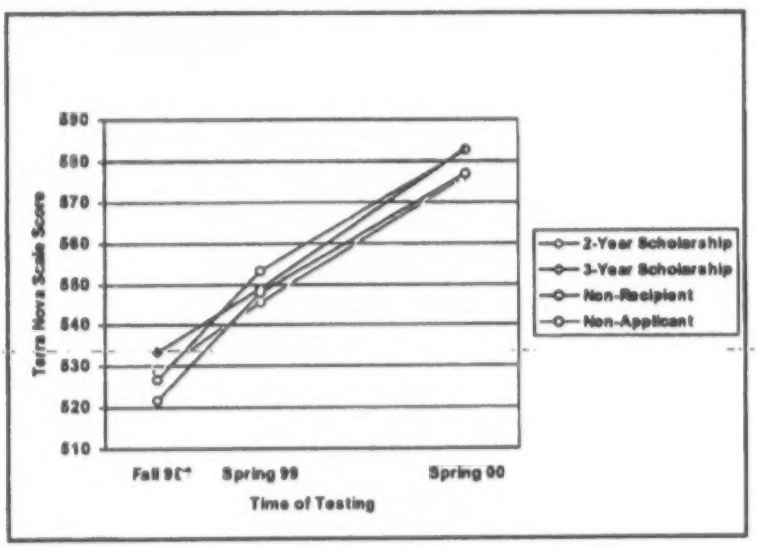


FIGURE 5. Total Achievement: 1998-00



BEST COPY AVAILABLE

Analyses first provided information about the effect of *time* on students' performance, independent of their scholarship status. Because we would hope that students' learning in schools is cumulative, we can assume that their test scores would increase as they experience more formal schooling and our data bear this out. Students' academic performance in each of the four areas improved significantly from the beginning of first grade to the end of first grade and from the end of first grade to the end of second grade. This was true for all students, independent of their scholarship status, and suggests that all students benefited significantly from their schooling, whether in private or public schools and whether they were one or two-year scholarship, applicant/non-recipient, or non-applicant students. This pattern of improvement is apparent in the positive slope of all four lines in the graphs.

A second result of the analytic technique for each area of the test was information about the effects of *group* membership on students' achievement, independent of time. With the exception of language, in which scholarship students who began the program in kindergarten had consistently higher scores than did non-applicants, no pattern of achievement difference was found based on scholarship status. Again, this largely disordinal pattern is observable in the four graphs presented above.

For the purposes of the present evaluation, the most important information provided by the analytic technique is an indication of the interaction of *group and time*. For example, if the Scholarship Program influences student achievement positively, a significant interaction of time and group would appear as more positive slopes (a greater angle of increase) in the lines representing scholarship student performance in the graphs. This would reflect a more rapid or greater rate of increase over time for scholarship students than for students in other groups. In each of the four achievement analyses, an interaction effect was found to be significant. However, the trend of this interaction was not consistent across the four measures, nor did it favor any one group. In general, students who entered the Scholarship Program as kindergartners (and thus

had participated for three years by the end of second grade) began first grade achieving at significantly higher levels than other students. And, public school students whose parents had never applied for a scholarship were often achieving at lower levels than other students when they began first grade.

By the end of first grade, earlier between group differences were reduced or eliminated. Non-applicants improved substantially in each of the four areas during their first grade year, and public school applicant/non-recipients improved similarly, though not as consistently or dramatically. Graphical presentation of students' achievement in each of the four areas indicates that the pattern of general improvement between students attending public schools (applicant/non-recipients and non-applicants) appears to differ from that of scholarship students. One year later, at the end of second grade, some differences again begin to emerge. By this point, scholarship students who began as kindergartners and applicant/non-recipients had improved more than either 2-year scholarship students (those who entered the program in first grade) or non-applicants.

The patterns of achievement gain over time and between groups are not clear, and data are only available for the first two years of students' formal schooling. As a result, it is much too early to draw definitive conclusions about the impact of the voucher program on students' achievement. As we have maintained throughout our work, and as research on vouchers in Milwaukee and other cities has indicated, the effects of schooling, and thus of the Scholarship Program, are cumulative and incremental. Further, research of other voucher programs indicates that the positive effects of choice are more likely as students move beyond the primary or early elementary grades.[5]

5. John F. Witte, *The Market Approach to Education: An Analysis of America's First Voucher Program* (Princeton, NJL Princetom University Press, 2000); W. G. Howell, P. J. Wolf, P. E. Peterson, and D. E. Campbell, "Test-Score Effects of School Vouchers in Dayton, Ohio, New York City, and Washington, DC: Evidence from Randomized Field Trials" (Paper presented at the annual meeting of the American Political Science Association, Washington, DC, September, 2000).

BEST COPY AVAILABLE

SUMMARY

Across the evaluation findings from these early years of students' experience in the Scholarship Program, there is little that is provocative. Scholarship students enter the program with somewhat different academic and demographic characteristics than students who attend public schools. Public school teachers are more likely to have completed graduate coursework, particularly up to the masters' level, than are private school teachers. Class size and teacher experience, at least for the samples of classrooms we examined, were quite similar between public and private schools. And student academic achievement, a factor watched closely by those on both sides of the voucher issue, presented no clear or consistent pattern than can be attributable to program participation.

The findings and the conclusions that we have attempted to draw are preliminary. They represent only the first of what are to be multiple sets of findings drawn on students over a period of several years. As a result, the findings are incomplete and understandably inconclusive. Yet, information can be gained about the ways in which a choice program, or schools generally, influence the early school experience of children.

Advocates and opponents of the Cleveland Scholarship Program will each find information from the present study useful and supportive of their position. However, the findings also suggest that what is often assumed ideologically to be true for all students, or all schools and teachers, or all voucher programs, may not necessarily be true at all. Only the continued generation of empirical evidence, collected in the most rigorous and neutral manner possible, can help policy makers, educators, and parents make informed and effective educational choices for children. It is this respect that the publicly-funded pilot program in Cleveland and the ongoing evaluation that it includes are critical to the continued discussion surrounding the voucher issue.